

Confidence Intervals for Estimators of Welfare Indices under Complex Sampling

by

Retha Kirchoff

Study Leaders:

Prof. T. de Wet and Dr. A. Neethling

Assignment presented in partial fulfillment of the requirements for the degree of Master of Commerce at Stellenbosch University.



UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

Department of Statistics and Actuarial Science,
Stellenbosch University.

February 2010

To my Mother.

You were always my greatest supporter.

Declaration

I, the undersigned, hereby declare that the work contained in this assignment is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Signature:
R. Kirchoff

Date:

Abstract

The aim of this study is to obtain estimates and confidence intervals for welfare indices under complex sampling. It begins by looking at sampling in general with specific focus on complex sampling and weighting. For the estimation of the welfare indices, two resampling techniques, viz. jackknife and bootstrap, are discussed. They are used for the estimation of bias and standard error under simple random sampling and complex sampling. Three confidence intervals are discussed, viz. standard (asymptotic), percentile and bootstrap- t . An overview of welfare indices and their estimation is given. The indices are categorized into measures of poverty and measures of inequality. Two Laeken indices, viz. at-risk-of-poverty and quintile share ratio, are included in the discussion. The study considers two poverty lines, namely an absolute poverty line based on percy (ratio of total household income to household size) and a relative poverty line based on equivalized income (ratio of total household income to equivalized household size). The data set used as surrogate population for the study is the Income and Expenditure survey 2005/2006 conducted by Statistics South Africa and details of it are provided and discussed. An analysis of simulation data from the surrogate population was carried out using techniques mentioned above and the results were graphed, tabulated and discussed. Two issues were considered, namely whether the design of the survey should be considered and whether resampling techniques provide reliable results, especially for confidence intervals. The results were a “mixed bag”. Overall, however, it was found that weighting showed promise in many cases, especially in the improvement of the coverage probabilities of the confidence intervals. It was also found that the bootstrap resampling technique was reliable (by looking at standard errors). Further research options are mentioned as possible solutions towards the mixed results.

Uittreksel

Die doel van die studie is die verkryging van beramings en vertrouensintervalle vir welvaartsmaatstawwe onder komplekse steekproefneming. 'n Algemene bespreking van steekproefneming word gedoen waar daar spesifiek op komplekse steekproefneming en weging gefokus word. Twee hersteekproefnemingstegnieke, nl. uitsnit (jackknife)- en skoenustersteekproefneming, word bespreek as metodes vir die beraming van die maatstawwe. Hierdie maatstawwe word gebruik vir sydigheidsberaming asook die beraming van standaardfoute in eenvoudige ewekansige steekproefneming asook komplekse steekproefneming. Drie vertrouensintervalle word bespreek, nl. die standaard (asimptotiese), die persentiel en die bootstrap- t vertrouensintervalle. Daar is ook 'n oorsigtelike bespreking oor welvaartsmaatstawwe en die beraming daarvan. Hierdie welvaartsmaatstawwe vorm twee kategorieë, nl. maatstawwe van armoede en maatstawwe van ongelykheid. Ook ingesluit by hierdie bespreking is die “at-risk-of-poverty” en “quintile share ratio” wat deel vorm van die Laekenindekse. Twee “armoedemaatlyne”, 'n absolute- en relatiewemaatlyn, word in hierdie studie gebruik. Die absolute armoedemaatlyn word gebaseer op “percy”, die verhouding van die totale huishoudingsinkomste tot die grootte van die huishouding, terwyl die relatiewe armoedemaatlyn gebaseer word op “equivalized income”, die verhouding van die totale huishoudingsinkomste tot die “equivalized” grootte van die huishouding. Die datastel wat as surrogaat populasie gedien het in hierdie studie is die Inkomste en Uitgawe opname van 2005/2006 wat deur Statistiek Suid-Afrika uitgevoer is. Inligting met betrekking tot hierdie opname word ook gegee. Gesimuleerde data vanuit die surrogaat populasie is geanaliseer deur middel van die hersteekproefnemingstegnieke wat genoem is. Die resultate van die simulاسie is deur middel van grafieke en tabelle aangedui en bespreek. Vanuit die simulاسie het twee vrae opgeduik, nl. of die ontwerp van 'n steekproef, dus weging, in ag geneem behoort te word en of die hersteekproefnemingstegnieke betroubare resultate lewer, veral in die geval van die

vertrouensintervalle. Die resultate wat verkry is, het baie gevarieer. Daar is egter bepaal dat weging in die algemeen belowende resultate opgelewer het vir baie van die gevalle, maar nie vir almal nie. Dit het veral die dekkingswaarskynlikhede van die vertrouensintervalle verbeter. Daar is ook bepaal, deur na die standaardfoute van die skoenuitberamers te kyk, dat die skoenuitsteekniek betroubare resultate gelever het. Verdere navorsingsmoontlikhede is genoem as potensiële verbeteringe op die gemengde resultate wat verkry is.

Acknowledgements

I would like to express my sincere gratitude to the following persons:

- My father for giving me the opportunity to study and get this far.
- My husband, André, for all his love, support and understanding without which I could not have achieved this.
- My study leaders, Prof. T. de Wet and Dr. A. Neethling, for all their patience, hard work, inspiration and guidance throughout the two years. Thank you for believing in me.

Soli Deo Gloria!

Contents

1	Introduction	22
1.1	The Problem Statement	22
1.2	Outline of the Study	23
2	Sampling	27
2.1	Probability and Non-probability Sampling	27
2.2	Probability Sampling Methods	28
2.2.1	Simple Random Sampling	28
2.2.2	Systematic Sampling	29
2.2.3	Stratified Random Sampling	30
2.2.4	Cluster Sampling	31
2.3	Complex Sampling	32
2.4	Weighting	34
2.4.1	Development of a Design Weight	36
2.4.2	Adjustment of Sample Weights for non-response	39
2.4.3	Adjustment of Sample Weights for Non-Coverage and Under-Representation	41
2.4.4	Post-Stratification and Cell-Weighting	42
2.4.5	Calibration Weighting	42
2.4.6	Integrated Weighting	45
2.5	Estimation	48
2.6	Conclusion	49
3	Resampling Techniques	52
3.1	The Jackknife	53
3.1.1	Estimation of Bias	54

3.1.2	Estimation of Standard Error	56
3.1.3	A comparison of the Jackknife method and Linearization Method	60
3.2	The Bootstrap	61
3.2.1	Estimation of Bias	62
3.2.2	Estimation of Standard Error	64
3.2.3	The choice of B	66
3.2.4	The choice of m_h	67
3.2.5	Conclusion	68
4	Confidence Intervals	69
4.1	The Standard (Asymptotic) Interval	69
4.2	The Percentile Interval	71
4.3	The Bootstrap- t Interval	73
4.4	Conclusion	79
5	Indicators of Poverty and Inequality	80
5.1	Measures of poverty	80
5.1.1	Headcount Index	82
5.1.2	Poverty Gap Index	83
5.1.3	Squared Poverty Gap Index	83
5.2	Measures of Inequality	84
5.2.1	Gini Coefficient of Inequality	85
5.2.2	Generalized Entropy Measures	87
5.2.3	Atkinson's Inequality Measures	88
5.3	Laeken Indicators	89
5.3.1	Definitions of Indicators	90
5.3.2	Estimation of Indicators	92
5.4	Conclusion	94
6	Description of the Data	95
6.1	Income and Expenditure Survey 2005/2006	95
6.1.1	Data Collection Methods	96
6.1.2	Response and Imputation of non-response	97
6.2	Survey Design	97
6.3	Weighting	98

6.4	Simulated Datasets	99
6.5	Conclusion	100
7	Analysis	101
7.1	Introduction	101
7.2	Discussion of Simulation Results	109
7.2.1	Estimators of Welfare Indices	109
7.2.2	Confidence Intervals for Estimators of Welfare Indices	133
7.3	Summary and Conclusions	162
8	Conclusions and Further Research	182
	Bibliography	186

List of Figures

4.1	Bootstrap- t Confidence Interval	75
5.1	Lorenz Curve	86
7.1	True Bias of Estimates of Poverty Measures based on Percy	111
7.2	True Bias of Estimates of Inequality Measures based on Percy	112
7.3	True MSE of Estimates of Poverty Measures based on Percy	112
7.4	True MSE of Estimates of Inequality Measures based on Percy	113
7.5	True Bias of Estimates of Poverty Measures based on Equivalised Income	114
7.6	True Bias of Estimates of Inequality Measures based on Equivalised Income	115
7.7	True Bias of Estimates of Inequality Measures, Equivalised Income: Quintile Share Ratio	116
7.8	True MSE of Estimates of Poverty Measures based on Equivalised Income	116
7.9	True MSE of Estimates of Inequality based on Equivalised Income	117
7.10	True MSE of Estimates of Inequality, Equivalised Income: Quintile Share Ratio	117
7.11	Difference between Bootstrap Bias and True Bias of Poverty Measures (Percy)	119
7.12	Difference between Bootstrap Bias and True Bias of Inequality Mea- sures (Percy)	120
7.13	Difference between Bootstrap MSE and True MSE of Poverty Mea- sures (Percy)	120
7.14	Difference between Bootstrap MSE and True MSE of Inequality Mea- sures (Percy)	121

7.15	Difference between Bootstrap Bias and True Bias of Poverty Measures (Equivalised Income)	123
7.16	Difference between Bootstrap Bias and True Bias of Inequality Measures (Equivalised Income)	123
7.17	Difference between Bootstrap Bias and True Bias of Inequality Measures (Equivalised Income): Quintile Share Ratio	124
7.18	Difference between Bootstrap MSE and True MSE of Poverty Measures (Equivalised Income)	124
7.19	Difference between Bootstrap MSE and True MSE of Inequality Measures (Equivalised Income)	125
7.20	Difference between Bootstrap MSE and True MSE of Inequality Measures (Equivalised Income): Quintile Share Ratio	125
7.21	% Relative Bias of a variance estimator of Poverty Measures (Percy) .	126
7.22	% Relative Bias of a variance estimator of Inequality Measures (Percy)	127
7.23	% Relative Bias of a variance estimator of Poverty Measures (Eq Income)	128
7.24	% Relative Bias of a variance estimator of Inequality Measures (Eq Income)	129
7.25	% Relative Bias of a variance estimator of Poverty Measures (Eq Income): Quintile Share Ratio	130
7.26	% Relative Bias of a MSE estimator of Poverty measures (Percy) . .	131
7.27	% Relative Bias of a MSE estimator of Inequality measures (Percy) .	131
7.28	% Relative Bias of a MSE estimator of Poverty measures (Eq Income)	132
7.29	% Relative Bias of a MSE estimator of Inequality measures (Eq Income)	132
7.30	% Relative Bias of a MSE estimator of Inequality measures (Eq Income), without Quintile Share Ratio	133
7.31	Comparison of 95% Confidence Intervals: No Weighting (Percy) . . .	134
7.32	Comparison of 95% Confidence Intervals: No Weighting (Eq Income)	135
7.33	Comparison of 95% Confidence Intervals: Design Weight (Percy) . . .	135
7.34	Comparison of 95% Confidence Intervals: Design Weight (Eq Income)	136
7.35	Comparison of 95% Confidence Intervals: RR_pp (Percy)	136
7.36	Comparison of 95% Confidence Intervals: RR_pp (Eq Income) . . .	137
7.37	Comparison of 95% Confidence Intervals: RR_ph (Percy)	137
7.38	Comparison of 95% Confidence Intervals: RR_ph (Eq Income) . . .	138

7.39	95% Standard Confidence Interval Length: Poverty Measures (Percy)	142
7.40	95% Standard Confidence Interval Length: Inequality Measures (Percy)	142
7.41	95% Standard Confidence Interval Length: Poverty Measures (Eq Income)	143
7.42	95% Standard Confidence Interval Length: Inequality Measures (Eq Income)	143
7.43	95% Standard Confidence Interval Length: \hat{I}_2 (Eq Income)	144
7.44	95% Percentile Confidence Interval Length: Poverty Measures (Percy)	144
7.45	95% Percentile Confidence Interval Length: Inequality Measures (Percy)	145
7.46	95% Percentile Confidence Interval Length: Poverty Measures (Eq Income)	145
7.47	95% Percentile Confidence Interval Length: Inequality Measures (Eq Income)	146
7.48	95% Percentile Confidence Interval Length: \hat{I}_2 (Eq Income)	146
7.49	95% Bootstrap- t Confidence Interval Length: Poverty Measures (Percy)	147
7.50	95% Bootstrap- t Confidence Interval Length: Inequality Measures (Percy)	148
7.51	95% Bootstrap- t Confidence Interval Length: Poverty Measures (Eq Income)	148
7.52	95% Bootstrap- t Confidence Interval Length: Inequality Measures (Eq Income)	149
7.53	95% Bootstrap- t Confidence Interval Length: \hat{I}_2 (Eq Income)	149
7.54	Comparison of Standardized Lengths of Confidence Intervals: Design Weight vs No Weight (Percy)	151
7.55	Comparison of Standardized Lengths of Confidence Intervals: Design Weight vs No Weight (Eq Income)	152
7.56	Comparison of Standardized Lengths of Confidence Intervals: RR_pp Weight vs No Weight (Percy)	153
7.57	Comparison of Standardized Lengths of Confidence Intervals: RR_pp Weight vs No Weight (Eq Income)	154
7.58	Comparison of Standardized Lengths of Confidence Intervals: RR_ph Weight vs No Weight (Percy)	155
7.59	Comparison of Standardized Lengths of Confidence Intervals: RR_ph Weight vs No Weight (Eq Income)	155

7.60	95% Standard Confidence Interval Shape: Poverty Measures (Percy) .	156
7.61	95% Standard Confidence Interval Shape: Inequality Measures (Percy)	157
7.62	95% Percentile Confidence Interval Shape: Poverty Measures (Percy)	157
7.63	95% Percentile Confidence Interval Shape: Inequality Measures (Percy)	158
7.64	95% Percentile Confidence Interval Shape: Poverty Measures (Eq In- come)	158
7.65	95% Percentile Confidence Interval Shape: Inequality Measures (Eq Income)	159
7.66	95% Bootstrap- t Confidence Interval Shape: Poverty Measures (Percy)	160
7.67	95% Bootstrap- t Confidence Interval Shape: Inequality Measures (Percy)	161
7.68	95% Bootstrap- t Confidence Interval Shape: Poverty Measures (Eq Income)	161
7.69	95% Bootstrap- t Confidence Interval Shape: Inequality Measures (Eq Income)	162

List of Tables

7.1	True values of welfare indices	109
7.2	Mean estimated values of welfare indices under different weighting methods: Percy	110
7.3	Mean estimated values of welfare indices under different weighting methods: Eq Income	110
7.4	True Bias and MSE for Welfare Indices under different weighting methods: Percy	111
7.5	True Bias and MSE for Welfare Indices under different weighting methods: Eq Income	114
7.6	Mean Estimated Bias and MSE for Welfare Indices under different weighting methods with the bootstrap resampling method: Percy . .	118
7.7	Mean Estimated Bias and MSE for Welfare Indices under different weighting methods with the bootstrap resampling method: Eq In- come	122
7.8	Non-coverage Probability of the 95% Standard Confidence Interval . .	139
7.9	Non-coverage Probability of the 95% Percentile Confidence Interval .	140
7.10	Non-coverage Probability of the 95% Bootstrap- t Confidence Interval	141
7.11	Summary of performance measures calculated for headcount based on the absolute poverty line	163
7.12	Summary of performance measures calculated for poverty gap based on the absolute poverty line	165
7.13	Summary of performance measures calculated for squared poverty gap based on the absolute poverty line	166
7.14	Summary of performance measures calculated for headcount based on the relative poverty line	167

7.15	Summary of performance measures calculated for poverty gap based on the relative poverty line	168
7.16	Summary of performance measures calculated for squared poverty gap based on the relative poverty line	169
7.17	Summary of performance measures calculated for AROP based on the relative poverty line	170
7.18	Summary of performance measures calculated for Gini coefficient based on percy	172
7.19	Summary of performance measures calculated for MLD based on percy	173
7.20	Summary of performance measures calculated for theil T based on percy	174
7.21	Summary of performance measures calculated for Gini based on equivalized income	175
7.22	Summary of performance measures calculated for MLD based on equivalized income	177
7.23	Summary of performance measures calculated for theil T based on equivalized income	178
7.24	Summary of performance measures calculated for QSR based on equivalized income	179

Notation

Some of the important notation used in this study is defined here for convenience.

- F : Unknown distribution of the data.
- \hat{F} : The empirical distribution used to estimate F .
- N : The population size.
- \hat{N} : The estimated population size.
- R : Number of bootstrap populations to be formed from the surrogate population.
- B : Number of bootstrap samples to be taken with replacement from each bootstrap population.
- z : Used to denote a poverty line.
- $percy$: Ratio of total household income to household size.
- $eqinc_i$: Equivalized income of person i , the ratio of total household income to equivalized household size.
- PSU: Primary sampling unit (enumerated areas (EA's) in this study).
- SSU: Secondary sampling unit (Households).
- USU: Ultimate sampling unit (Persons in a household).
- n : The sample size.
- h : The stratum index.

- j : The psu index.
- i : The usu index.
- θ : The population parameter of interest.
- $\hat{\theta}$: The estimator of the parameter of interest.
- \mathbf{y} : A vector of values (y_1, \dots, y_n) , denoting the sample/variable of interest.
- y_{hji} : An observation i in the j th psu in the h th stratum.
- p_i : The inclusion probability of the i th sampling unit (in weighting).
- w_i : The weight of the i th sampling unit (in weighting).
- p_j : Is the probability that the j th psu is selected.
- A_j : The measure of size of the j th psu.
- M_j : The number of households (ssu's) in the j th psu.
- m_j : The number of ssu's selected from each sampled psu.
- w_{ji} : The design weight of the i th ssu in the j th psu.
- p_{ji} : The inclusion probability of the i th ssu in the j th psu.
- $p_{i|j}$: The conditional inclusion probability of the i th ssu in the j th psu.
- H : The number of strata.
- N_h : The number of psu's in the h th stratum.
- n_h : The number of psu's sampled from the h th stratum.
- p_{hji} : The inclusion probability of the i th sampling unit in the j th psu of the h th stratum.
- w_{hji} : The weight of the i th sampling unit in the j th psu of the h th stratum.
- r_{hj} : Response rate in the j th psu.
- A_{hj} : The measure of size of the j th psu in the h th stratum.

- M_{hj} : The number of households (ssu's) in the j th psu in the h th stratum.
- m_{hj} : The number of ssu's selected from the j th sampled psu in the h th stratum.
- w_{hji} : The final weight for the i th ssu in the j th psu in stratum h .
- w_{jik} : The weight from person k in household i in psu j .
- H_k : The size of the household to which person k belongs.
- $\hat{\theta}_{(i)}$: Jackknife estimator of the parameter of interest under a simple random sample (SRS).
- $\tilde{\theta}_{JK}$: Average of the jackknife estimates.
- $\hat{\theta}_{(hj)}$: Jackknife estimator of θ with the j th psu of the h th stratum deleted under a complex sample (CS).
- $w_{i(hj)}$: Jackknife adjusted weight used in the calculation of $\hat{\theta}_{(hj)}$.
- $\widehat{bias}_{JK}(\hat{\theta})$: Jackknife estimate of bias of $\hat{\theta}$.
- $\hat{V}_{JK}(\hat{\theta})$: Jackknife estimate of variance of $\hat{\theta}$.
- $\hat{\theta}^r$: The estimator of the parameter of interest for the r th jackknife population.
- $\hat{\theta}_{(hj)}^r$: The jackknife estimator obtained after the j th psu of the h th stratum in the r th population has been deleted.
- $\hat{V}_{JK}(\hat{\theta}^r)$: The jackknife estimated variance of $\hat{\theta}^r$.
- \mathbf{y}_b^* : A bootstrap sample.
- $\hat{\theta}_b^*$: A bootstrap replicate calculated on the bootstrap sample.
- $\hat{\theta}_{b_r}^*$: The bootstrap replicate calculated on the b th bootstrap sample of the r th bootstrap population.
- $\tilde{\theta}^*$: The average of the bootstrap replicates.
- $\widehat{bias}_B(\hat{\theta})$: Bootstrap estimate of the bias of $\hat{\theta}$.

- w_{hji}^* : The bootstrap weight used in the calculation of the bootstrap replicate under CS.
- m_{hj}^* : The number of times the j th psu occurred in the bootstrap sample.
- $\widehat{bias}_B(\hat{\theta}_r^*)$: The estimated bias of the r th bootstrap population.
- $\hat{V}_B(\hat{\theta})$: Bootstrap estimated variance of $\hat{\theta}$.
- $\hat{V}_B(\hat{\theta}_r^*)$: The bootstrap estimated variance of $\hat{\theta}_r^*$.
- P : Unknown probability model of the data.
- \hat{P} : Estimated probability model.
- \hat{G} : The cumulative distribution function of $\hat{\theta}^*$.
- $\hat{\theta}_\alpha^*$: $100 \cdot \alpha$ th bootstrap distribution percentile.
- $\hat{\theta}_{1-\alpha}^*$: $100 \cdot (1 - \alpha)$ th bootstrap distribution percentile.
- $\hat{\theta}_{(b)}^*$: The b th smallest bootstrap replicate.
- t_b^* : The bootstrap t-statistic.
- \widehat{se}_b^* : The estimated standard error of $\hat{\theta}_b^*$ for the bootstrap sample \mathbf{y}_b^* .
- $t_{(b)}^*$: The b th smallest bootstrap t-statistic.
- $V_{JK}(\hat{\theta}_b^*)$: The jackknife estimate of variance of the bootstrap replicate, $\hat{\theta}_b^*$.
- P_0 : Headcount index under SRS.
- P_H : Headcount index under CS.
- P_1 : Poverty gap index under SRS and CS.
- P_2 : Squared poverty gap under SRS and CS.
- $Gini$: Gini coefficient under SRS and CS.
- $GE(\alpha)$: General formula for the generalized entropy measures where α represents the weight associated with the distance between incomes at different levels of the income distribution.

- $GE(0)$: Theil's T index under SRS.
- $GE(1)$: Mean log deviation under SRS.
- $GE_w(0)$: Theil's T index under CS.
- $GE_w(1)$: Mean log deviation under CS.
- I_1 : At-risk-of-poverty index.
- I_2 : Quintile share ratio index.
- q_α : The $\alpha - th$ (sample) quantile.
- \hat{I}_{1e} : The at-risk-of-poverty threshold.
- \hat{I}_1 : Estimated at-risk-of-poverty.
- \hat{I}_2 : Estimated quintile share ratio.
- $bias(\hat{\theta})$: The true bias of an estimator.
- $MSE(\hat{\theta})$: The true MSE of an estimator.
- $\widehat{bias}_B(\hat{\theta}_r)$: The bootstrap estimated bias of $\hat{\theta}_r$.
- $\widehat{bias}_B(\hat{\theta})$: The overall bootstrap estimate of bias of $\hat{\theta}$.
- $Dev_{bias}(\hat{\theta})$: The difference between the bootstrap estimate of bias and the true bias.
- $\widehat{MSE}_B(\hat{\theta}_r)$: The bootstrap estimate of MSE of $\hat{\theta}_r$.
- $\widehat{MSE}_B(\hat{\theta})$: The overall bootstrap estimate of MSE of $\hat{\theta}$.
- $Dev_{MSE}(\hat{\theta})$: The difference between the bootstrap estimate of MSE and the true MSE.
- $\%RelBias(\widehat{MSE}(\hat{\theta}))$: The percentage relative bias of the estimated MSE of $\hat{\theta}$ with respect to $MSE(\hat{\theta})$.

- $\%RelBias\left(\widehat{var}\left(\hat{\theta}\right)\right)$: The percentage relative bias of the estimated variance of $\hat{\theta}$ with respect to $var\left(\hat{\theta}\right)$.
- l_r : The length of the confidence interval based on the r th bootstrap population.
- $\hat{\gamma}_r$: The shape of the confidence interval based on the r th bootstrap population.

Chapter 1

Introduction

1.1 The Problem Statement

When a survey (sample) is conducted, a sample of units is drawn from a target population and this sample is then used to obtain estimates of population parameters. This “simple” procedure gives rise to the question of how “good”, or how accurate, these estimators are.

Before considering the accuracy of the estimators, it is important to take into account the sample design, noncoverage and non-response. Failure to do so may result in biased estimates. The assigning of a design weight to each sample unit can result in a reduction in bias. The design weights are then adjusted for non-response and/or noncoverage and finally, the adjusted weights are used for estimation of population parameters.

A common measure of accuracy is the standard error, the square root of the variance, of an estimator. It is known that very few theoretical formulae exist for calculating standard errors under more complex sample designs and estimators. This obstacle is overcome by making use of resampling techniques such as the Bootstrap and Jackknife to estimate the standard errors. They provide a simple, robust approach to the estimation of sampling standard errors as well as the conducting of significance tests for survey data. These methods consist of estimating the standard errors of the estimators of the population parameters by taking subsamples from the obtained sample and calculating the estimator for each subsample. The variability between the calculated estimators is then used to estimate the variance of the initial estimator and the estimated standard error is obtained as the square root of the

estimated variance. Along with standard error, one should consider the bias of an estimator. This refers to how far the average statistic lies from the parameter it is estimating. Ideally one would desire an unbiased estimator where the expected value of the estimator equals the parameter it is estimating. Since this is not a common occurrence, bias is estimated by means of the resampling techniques as the difference between the average of the estimators calculated on each subsample and the parameter of interest. The squared standard error and the squared bias combine to give the mean squared error of an estimator, another equally important measure of accuracy.

An important part of any analysis is constructing confidence intervals to determine the accuracy of estimators. They combine point estimation and hypothesis testing into a single statement that gives more information than a point estimate or a standard error individually. The standard errors as described above can be used to form approximate confidence intervals for the population parameters of interest.

This study has three main objectives. Firstly to estimate the standard errors of estimators of population parameters by means of resampling. The resampling techniques that will be considered are the Bootstrap and Jackknife, the population referred to is the Income and Expenditure Survey 2005/2006 conducted by Statistics South Africa and the parameters of interest are the poverty and inequality measures that will be discussed. The second objective is to construct confidence intervals for the measures of poverty and inequality by means of the resampling techniques and the last objective is to determine the accuracy of these confidence intervals.

1.2 Outline of the Study

The study consists of seven chapters. In the next chapter sampling is discussed in general. Sections 2.1 and 2.2 discuss the difference between probability sampling and non-probability sampling as well as different probability sampling methods. In section 2.3 a definition of a complex sample is given along with a description of a common complex sample design. Since the population that will be considered in the simulation is based on a complex sample, it is important to form a good understanding of the concept. The potential hassles faced in complex samples are survey weights, non-response and the level of error associated with estimates due to the sample design. Thus, section 2.4 looks at weighting, an intricate part of complex

sampling that leads to wrong conclusions when not carried out correctly.

The design weight is used to deal with the effect of stratification and clustering on estimates. Weighting combines three stages, namely adjustment for unequal inclusion probabilities, non-response and adjustment of weighted estimates to match known population totals. A close look is taken at all three stages mentioned. The final stage of weighting consists of methods such as calibration and integrated weighting. Calibration makes use of auxiliary information in the form of known population totals and thus produces an adjusted set of weights called the calibration weights. Since calibration results in different weights for each household member, integrated weighting was developed to achieve a single representative set of weights that could be used for the estimation of both person and household characteristics. Finally, estimation with weights is discussed.

Chapter three contains a general discussion of resampling techniques and their application in the estimation of standard error and bias. The chapter appropriately starts with the jackknife method. It predates the bootstrap as a method of estimating standard error and bias. In a simple random sample of size n , the jackknife leaves out one unit at a time and calculates the estimator on the remaining $n - 1$ units. This results in n estimators and estimates the standard error of the estimator as the average squared difference between the jackknife estimator and the estimator on the original sample multiplied with a factor $n - 1$. It estimates the bias of an estimator as the difference between the average of the jackknife estimators and the estimator on the original sample multiplied by the factor $n - 1$. In a complex sample the jackknife method is applied independently in each stratum. It leaves out one primary sampling unit (psu) at a time and calculates an adjusted jackknife weight that is used in the calculation of the jackknife estimator. This weight adjustment is done to ensure that the weights sum to the correct stratum total. Finally, the standard error and bias is estimated in the same way as for the simple random sample, except that it sums over the strata. Refer to equation (3.16) for the jackknife estimate of standard error in complex sampling.

Section 3.2 discusses the bootstrap method. It was introduced in 1979 as a computer intensive approach to estimating the variance, or standard error, of an estimator and has the pleasing property of being able to produce an estimate regardless of the mathematical complexity of the estimator. In a simple random sample of size n it proceeds by drawing a with replacement sample of size n and calculating the

bootstrap replicate with the same form as the estimator. This is repeated a large number of times, B , and the variability between the bootstrap replicates is used to estimate the variance of the estimator. The bias is estimated as the difference between the average of the bootstrap replicates and the estimator based on the original sample. In a complex sample the bootstrap method is applied independently in each stratum. A with replacement sample of psu's is selected from each stratum and an adjusted bootstrap weight is calculated for the same reasons as given previously. These bootstrap weights are used to calculate the bootstrap replicates and the bootstrap estimate of variance is calculated as the variability between these bootstrap replicates. The bias is estimated in a similar way.

A brief discussion is given on the choice of the size of B . It refers to a conditional and unconditional coefficient of variation approach. The conditional coefficient of variation takes only resampling variability into account whilst the unconditional coefficient of variation takes both sampling and resampling variability into account. These approaches lead to very different choices of the size of B .

Confidence intervals are considered in chapter four. The combination of a point estimate and an interval estimate gives the “best guess” of the population parameter as well as how far that “guess” may be from the actual value of the parameter and hence a proper understanding of confidence intervals is necessary. In section 4.2 a brief overview of the standard (asymptotic) interval is given. This is followed by a discussion on the percentile interval in section 4.3. It is the simplest way of approximating a confidence interval through the application of the bootstrap. It merely uses the sorted bootstrap replicates at specific points as the upper and lower bounds of the interval. This method has many desirable properties and pleasing advantages that are outlined in that section. The percentile interval will be obtained in exactly the same way in a complex sample, but independently for each stratum and following the same bootstrap procedure as outlined in chapter three.

The bootstrap- t interval is discussed in section 4.4. This method makes use of a nested bootstrap to obtain the interval. The first level of bootstrap replicates are calculated followed by a second level of bootstrapping to estimate the standard error of these bootstrap replicates. The procedure is graphically explained in figure 4.1. Due to the computational demand of this method, the standard errors of the bootstrap replicates will be estimated with the jackknife method for standard error estimation carried out on the second level. The bootstrap- t interval will be

calculated independently in each stratum in a complex sample, but applying the jackknife weights on the second level to obtain the estimated standard errors of the bootstrap replicates. The properties, advantages and disadvantages of this interval are also discussed in the section.

Chapter five gives an overview of poverty and inequality measures. The poverty measures that will be considered, namely headcount index, poverty gap index and squared poverty gap, are defined as well as how to estimate them. These are followed by the definition and estimation of the inequality measures, namely Gini coefficient, generalized entropy measures and Atkinson's measures. Finally, the Laeken indicators are also defined.

Chapters six and seven are respectively the data description and analysis chapters. A description of the Income and Expenditure Survey 2005/2006 is given along with how the data was collected, how non-response was taken care of, the design of the survey and the weighting used. This is followed by a description of the smaller datasets that were used in the simulation study. Chapter seven examines the results obtained from the Monte Carlo simulation study. Section 7.1 gives an introduction to the chapter with discussions on what was done during the simulation study and what measures of accuracy were considered for both the estimators of the welfare indices as well as the confidence intervals. Section 7.2 contains the actual discussion of the results and the chapter is concluded with the results of the simulation being summarized in tabular form. Concluding remarks on each table as well as overall are given.

Chapter 8 summarises the findings in this study and contains suggestions for further research that could or should be done.

Chapter 2

Sampling

2.1 Probability and Non-probability Sampling

A wide variety of methods are available to collect information with regard to a particular study. The choice of collection technique, dependent on the study at hand, can either fall under the group of probability methods or nonprobability methods.

In probability sampling each population unit has a known probability of being selected for the sample [18]. This selection probability is greater than zero and can be accurately determined. This makes it possible to produce more unbiased estimates of population quantities as well as estimation of the standard errors of the estimators. Sampling methods included under probability sampling are:

1. Simple Random Sampling,
2. Systematic Sampling,
3. Stratified Sampling, and
4. Cluster Sampling.

Nonprobability sampling methods are characterized by some population units not having any chance of being selected for the sample. Alternatively, the selection probability of these population units can not be determined. Selection by these methods occurs through subjective evaluation, there are no probability techniques involved. In the absence of probability methods the survey estimates will be biased and more importantly, the extent of the biases will be unknown. Due to the nonrandom selection of units the sampling errors can not be estimated, which place limits on the

amount of population information that can be provided by the sample. Sampling methods included under nonprobability sampling are:

1. Convenient Sampling,
2. Quota Sampling, and
3. Purposive Sampling.

To conclude, only probability sampling can be used for inference since these are the only sampling methods for which selection probabilities, that are needed for the calculation of estimates and standard errors, can be determined.

2.2 Probability Sampling Methods

2.2.1 Simple Random Sampling

A simple random sample (SRS) of size n is the most common form of probability sampling where every possible subset of n population units has the same probability of being selected for the sample [18]. The accuracy of the results can easily be estimated since the variance between individual results within the sample is a good indicator of the population variance. There are two ways of taking a SRS:

1. Simple Random Sampling With Replacement (SRSWR)

This method can be seen as taking n independent samples of size one from the population of N units. The first sample unit is drawn with probability equal to $\frac{1}{N}$. This sampled unit is then placed back into the population. The second sample unit is drawn with the same probability of $\frac{1}{N}$. The procedure is repeated until all n units for the sample have been drawn. Thus, certain population units can appear more than once in the sample [18].

2. Simple Random Sampling Without Replacement (SRSWOR)

This is the preferred method of sampling since having the same population unit appear more than once in a sample provides no additional information with regard to the population. Simple random samples selected without replacement are selected in such a way that every possible subset of n distinct units in the population has the same probability of being selected. There are

$\binom{N}{n}$ possible samples of size n that can be drawn from the population which are each equally likely to be selected. Let $\mathbf{y} = \{y_1, \dots, y_n\}$ denote the sample. The probability of \mathbf{y} is given by

$$P(\mathbf{y}) = \frac{1}{\binom{N}{n}},$$

thus the probability of a unit to appear in \mathbf{y} is $\frac{n}{N}$ [18].

SRS can be vulnerable to sampling error due to the randomness of the selection process that could produce a sample that is not a good representation of the population. It should be noted that a SRS is always an equal probability sample (EPS), but an EPS is not always a SRS.

2.2.2 Systematic Sampling

Systematic Sampling (SS) can be used, instead of an SRS, when there is no list of the population units or when the population has been ordered according to some ordering scheme [18]. Firstly determine the selection interval defined as

$$k = \frac{N}{n},$$

and then select a random starting point, R , where R is a random number between 1 and k . This results in a sample of the form

$$S = \{R, R + k, R + 2k, \dots, R + (n - 1)k\}.$$

SS forms part of probability sampling as long as it makes use of a random starting point. It should be noted that it is not the same as a SRS since it does not possess the property that every possible group of n units has the same probability of being the sample. Although, if the population is in random order it will be much like a SRS. Its results can thus be compared to the results of a SRS and SRS methods can be used in the analysis of a sample selected with SS [18].

On the other hand, if periodicity exists in the list of population units, SS does not necessarily produce a representative sample since it is very vulnerable to periodicity [18]. If periodicity is present and the period is a multiple or factor of the interval

used, k , the sample is likely to be unrepresentative of the overall population. This will make the scheme less accurate than a SRS. To illustrate, let us assume that the population list alternates male and female names. Depending on whether k is an even or an odd number, the selected sample will either contain only male names or only female names.

2.2.3 Stratified Random Sampling

In stratified random sampling the population is divided into subgroups called strata and each population element belongs to only one stratum. A SRS is then taken independently from each stratum. The ideal is to have population units in the same stratum that are similar to each other such that the variation within a stratum is smaller than the variation between different strata. When this maximum between-strata variation and minimum within-strata variation is achieved, then a stratified random sample will give estimates with smaller standard error than estimates under a SRS [18].

Stratified random sampling is preferred to SRS since it has a smaller chance of producing a nonrepresentative sample of the population [18]. By stratifying the population into important/meaningful population subgroups we ensure proper representation of these subgroups without increasing the bias of the selection process. Stratified random samples are also easy to administer and could result in lower survey cost relative to a SRS [18].

There are two basic rules underlying stratified sampling. Firstly, a minimum of two units must be chosen per stratum to be able to calculate sampling error. Units in the same stratum tend to be more similar, homogeneous, resulting in less variability per stratum. Secondly, each stratum should differ substantially from the other. Thus, one would gain the most with heterogeneity between strata and homogeneity within strata. Increased reliability of estimates is gained with maximum heterogeneity between strata since this would result in the smallest standard error, i.e. best precision, of the estimates.

There are several benefits to using stratified sampling. The division of the population into distinct, independent strata makes it possible to draw inferences about specific subgroups that could be lost in more general random sampling. There is the increased efficiency of estimates as previously mentioned in comparison to a SRS. Finally, different sampling methods can be applied in different strata due to inde-

pendence which makes it possible to use a sampling method that is more appropriate for that particular subgroup than for another.

2.2.4 Cluster Sampling

Cluster sampling divides the population into larger subgroups called clusters. A sample of clusters is then drawn of which all, or some, of the units in each cluster are subsampled. The clusters are termed primary sampling units (psu) and the units subsampled from each selected cluster are termed secondary sampling units (ssu). When all the ssu's are sampled in a cluster, it is called one-stage cluster sampling. When a subsample of ssu's are taken from each selected psu it is called two-stage cluster sampling [18].

Benefits of cluster sampling, apart from cost effectiveness, are that there is no need for a complete sampling frame of population units but only of the clusters. The disadvantages include reliance of the accuracy of estimates on the clusters chosen. If the choice of clusters is unrepresentative, this will lead to very inaccurate estimates. The decrease in reliability of estimates is due to ssu's in a psu being more homogeneous than in a SRS. To compensate for this there has to be an increase in the sample size.

There are two ways in which cluster sampling differs from stratified sampling. In stratified sampling a sample of units is taken from each stratum which ensures the inclusion of all strata in the sample. In cluster sampling, only a selection of clusters is made, thus they represent those clusters that were not sampled. Strata are chosen to be homogeneous within and heterogeneous between strata. The opposite is true for clusters. The clusters need to be internally as heterogeneous as possible for the sake of precision. The degree of clustering in the sample speaks to its reliability as a representation of the population.

Clusters can be drawn in two ways [19]:

1. with equal probability irrespective of the number of population elements within each cluster, or
2. with probability proportionate to some meaningful measure of size (MOS) of clusters.

To illustrate the difference between these, assume a survey is undertaken that wishes to estimate the number of learners in grade 12 that take Mathematics as a subject.

Equal probability sampling of clusters would occur when a sample of secondary schools is taken, irrespective of the number of grade 12 learners in each school. Here the schools would be the psu's and each school will have a probability of

$$p = \frac{n}{N} \quad (2.1)$$

to be included in the sample where n is the number of psu's being drawn for the sample and N is the number of psu's in the population [19].

Probability proportionate to some measure of size would occur when the sample of secondary schools is taken according to the number of grade 12 learners in each school. In this case, let A_j be the measure of size of the j th psu where

$$\sum_{j=1}^N A_j$$

is the total measure of size of the population. The inclusion probability of the j th psu will be

$$p_j = n \cdot \frac{A_j}{\sum_{j=1}^N A_j}, \quad (2.2)$$

where n is the number of psu's drawn for the sample and N is the number of psu's in the population [19].

2.3 Complex Sampling

A complex sample (CS) is defined as a stratified multistage cluster sample. The population is divided into strata to ensure the sample has effective representation of the population. Within each stratum smaller groups are formed called clusters. These are referred to as primary sampling units, hereafter referred to as psu's. A predetermined number of psu's are then drawn from each stratum with probability sampling. The population elements in each drawn psu are then grouped into smaller clusters called secondary sampling units, hereafter referred to as ssu's. A predetermined number of ssu's are then drawn from each psu. This procedure is repeated until the population elements, or groups of population elements, termed ultimate sampling units (usu's), are drawn. Note that the psu's can be stratified before ssu's

are drawn and that at least two psu's per stratum have to be drawn to be able to estimate the variance.

Advantages of this method of sampling are the benefit of designing the sample step-by-step and the fact that complex surveys are more economically and practically viable. If a straightforward simple random sample (SRS) of households in South Africa were to be selected, it could result in fieldworkers having to travel from one household in rural Western Cape to, for example, an urban household in Mpumalanga incurring enormous travel costs. Also, no complete sampling frame of population elements is required for CS, only a sampling frame of psu's. Since the design can be controlled more effectively, a better representative sample can be designed [21].

The disadvantage is that complex sampling is generally less efficient than simple random sampling which results in estimates with lower precision relative to a SRS for a fixed sample size [21]. Here we have to consider the design effect (*deff*) to determine the effect of using a non-SRS design. Let $\hat{\theta}$ be the estimator of the parameter of interest. The design effect is defined as the combined effect of stratification and clustering on the variance of an estimate, $V(\hat{\theta}|CS)$, in comparison to the variance of the same estimate obtained under simple random sampling, $V(\hat{\theta}|SRS)$. It provides a measure of the precision gained or lost due to the use of complex sampling rather than simple random sampling [18]. Thus

$$deff = \frac{V(\hat{\theta}|CS)}{V(\hat{\theta}|SRS)}, \quad (2.3)$$

where $\hat{\theta}$ is the estimator and (2.3) expresses the ratio of the variance of the estimator obtained under complex sampling to the variance of the estimator under simple random sampling.

Stratification generally results in smaller variances of estimates, or alternatively, results in estimates with better precision due to the fact that the population is divided into homogeneous groups (strata). By contrast, cluster sampling usually results in a loss of precision, because elements in a cluster are more similar than the stratum to which they belong. The total design effect is then dependent on whether more precision is gained through the stratification than lost through the clustering [18].

To improve the efficiency of complex sampling a larger sample is required than

a SRS. Although this is the case, complex sampling is still more convenient and has lower cost per unit than simple random sampling. This may result in obtaining the same precision through complex sampling as through simple random sampling at a lower cost even if a larger sample is needed [18]. Potential hassles one is faced with in the analysis of complex sample data are [27]:

1. Survey weights.
2. Non-response.
3. Sample design that impacts on the level of error associated with estimates obtained from the data.

In the next section, weighting will be discussed along with calibration and integrated weighting.

2.4 Weighting

Weighting is a method used to deal with the effects that stratification and clustering have on estimates. The sampling weight of an observation is defined as the reciprocal of the probability that the observation is selected to be in the sample. In a SRS, let p_i denote the inclusion probability and w_i denote the weight of the i th sampling unit. Then,

$$w_i = \frac{1}{p_i}, i = 1, \dots, n, . \quad (2.4)$$

In a CS, let p_{hji} denote the inclusion probability and w_{hji} denote the weight of the i th sampling unit in the j th psu of the h th stratum. Then,

$$w_{hji} = \frac{1}{p_{hji}}, h = 1, \dots, H, j = 1, \dots, n_h, i = 1, \dots, n_{hj}. \quad (2.5)$$

A weight can be thought of as the number of population units represented by the corresponding sampling unit. It is known that the sum of the sampling weights provide an unbiased estimate of the population size [18]. In many situations the sample fraction, i.e. the ratio of the sample size to the total number of population units, may be varied by stratum resulting in disproportional allocation. As a result the correct weights have to be used to ensure that the weights weigh up to the correct

stratum total. In general, data should be weighted if the sample is designed in such a way that each unit does not have the same inclusion probability. For instance, if households are selected with equal probabilities and only one person from every selected household is chosen to be interviewed, a person from a large household has a smaller probability of being interviewed, resulting in unequal inclusion probabilities.

Sampling weights are applied to the data before any analysis is done to ensure that the sample records represent the target population as closely as possible. These weights combine three stages:

1. It adjusts for effect of different sampling rates applied to different population subgroups (unequal inclusion probabilities).
2. Adjust for differences in the rate of survey non-response among different demographic subgroups in the population.
3. Weight adjustment of the sums of the sampling weights of the various subgroups so that these agree with the size of the subpopulations.

The first two stages try to remove any bias that may occur as a result of differences in selection probabilities produced by the sampling design as well as reduce possible bias due to differences in non-response. The third stage is introduced to establish consistency with known population counts and to increase precision [27]. All information needed to construct point estimates are contained in the sampling weights, but the weights give no information on the construction of the standard errors of the estimates. Thus, sampling weights alone do not enable one to do inference. For inference one needs more information on the sampling design, since variances of estimators are dependent on the probability that any unit pair is selected and requires more knowledge of the sampling design used than given by weights alone [18].

If large weights appear in the dataset, they could be truncated so that a single observation does not contribute significantly to the overall estimate. Although this introduces some bias to the estimators, it also reduces the mean squared error (MSE). During estimation there is a constant bias-variance trade-off [18].

Various techniques exist that can be used to adjust the sampling weights depending on the availability of auxiliary information [21]. Such techniques include

1. cell-weighting and post-stratification;
2. calibration; and

3. integrated weighting.

The calibration and integrated weighting techniques will be considered in sections 2.4.4 and 2.4.5.

2.4.1 Development of a Design Weight

Weights are developed in different stages and are assigned to respondents so that the sample represents the population as closely as possible. In the first stage of weighting a base weight, called the design weight, is assigned to each sampled element [20]. The design weights are defined in (2.4) and (2.5) and should reflect the selection probability at each level in the complex sample design [23].

The second stage of weighting sees the adjustment of the design weights to compensate for non-response. This is done by increasing the design weight of respondents in each sample weighting cell by the inverse of the response rate, as will be discussed in section 2.4.2 [20].

The final stage of weighting involves the use of auxiliary information to adjust the weighted estimates so that they match known population totals of certain key variables. Auxiliary information refers to additional information known about the finite population. This adjustment is made to reduce variances of estimators.

For the calculation of the design weight in the first stage, let us firstly consider a two-stage design where p_j is the probability that the j th psu is selected. Define the following notation:

- N : The number of psu's in the population;
- n : The number of psu's in the sample;
- A_j : The measure of size of the j th psu;
- M_j : The number of households (ssu's) in the j th psu; and
- m_j : The number of ssu's selected from the j th psu.

If the psu's are selected with equal probability, the inclusion probability of a psu is the fraction $\frac{n}{N}$ as given in equation (2.1) and the conditional inclusion probability of a ssu in the j th sampled psu, is $\frac{m_j}{M_j}$ if they are also drawn with equal probability

within the selected psu [23]. Thus, the inclusion probability of the i th ssu in the j th psu is given by

$$p_{ji} = \frac{n}{N} \cdot \frac{m_j}{M_j},$$

and the overall design weight equals

$$w_{ji} = \frac{1}{p_{ji}} = \frac{N}{n} \cdot \frac{M_j}{m_j}.$$

If the psu's are selected with some measure of size (MOS), the inclusion probability of the j th psu is given by

$$p_j = n \cdot \frac{A_j}{\sum_j A_j},$$

and the conditional inclusion probability of the i th ssu in the j th psu is given by

$$p_{i|j} = \frac{m_j}{M_j}.$$

Thus, the total inclusion probability of the i th ssu in the j th psu is

$$p_{ji} = n \cdot \frac{A_j}{\sum_j A_j} \times \frac{m_j}{M_j},$$

and the weight of that ssu will be

$$w_{ji} = \frac{1}{n} \frac{\sum A_j}{A_j} \times \frac{M_j}{m_j}. \quad (2.6)$$

When the number of ssu's in the j th psu is used as MOS (i.e. $A_j = M_j$), the inclusion probability of the i th ssu simplifies to

$$p_{ji} = n \cdot \frac{m_j}{\sum_j A_j},$$

and the weight simplifies to

$$w_{ji} = \frac{1}{n} \frac{\sum A_j}{m_j}.$$

Consider the following notation when the two-stage design considered above is extended to a stratified multistage sample design:

- H : The number of strata;
- N_h : The number of psu's in the h th stratum;
- n_h : The number of psu's sampled from the h th stratum;
- A_{hj} : The measure of size of the j th psu in the h th stratum;
- M_{hj} : The number of households (ssu's) in the j th psu in the h th stratum; and
- m_{hj} : The number of ssu's selected from the j th sampled psu in the h th stratum.

Then the final weight for the i th ssu in stratum h , according to equation (2.6), will be

$$w_{hji} = \frac{1}{n_h} \frac{\sum A_{hj}}{A_{hj}} \times \frac{M_{hj}}{m_{hj}}. \quad (2.7)$$

These weights are calculated independently for each stratum.

Self-weighting Samples

A self-weighting sample is obtained when the sampling weights of all sampled units are the same. Units in higher stages of the multistage design are selected with varying probabilities in order to obtain an efficient sample.

Self-weighting samples are rarely obtained in household surveys where one person per household is selected. The reasons being that, firstly, although the sample is designed to be self-weighting on household level, unequal probabilities are achieved because one member is selected to be interviewed from sampled households which are not equally sized. In this case, the weight from person k in household i in psu j is:

$$w_{jik} = \frac{N}{n} \cdot \frac{M_j}{m_j} \cdot H_k,$$

where H_k is the size of the household to which person k belongs.

Secondly, imperfections such as non-response and non-coverage almost always occur in sampling. Thirdly, when stratified multistage cluster sampling is used, disproportional allocation is used to increase the sample size of smaller strata to meet precision requirements [23].

The advantages underlying a self-weighting sample are so attractive that it should always be the goal of any sampling design to produce a self-weighting sample. For instance, estimates can be obtained from unweighted data and, if necessary, the results can be inflated by a constant factor to obtain appropriate estimates of population parameters [23].

2.4.2 Adjustment of Sample Weights for non-response

The presence of non-response will have an effect on any well designed survey. Thus, “the best way to deal with non-response is to prevent it [18]”. Two types of non-response occur:

1. Unit non-response:

The entire observation unit is missing.

2. Item non-response:

Some parts of the observation unit are missing.

Due to the existence of non-response, the sampled units rarely provide all the information required. Any regular differences between respondents and non-respondents will lead to bias estimates based only on the respondents. Thus, the non-response should be kept as low as possible [23]. There are many ways to deal with non-response, such as

- Prevention, where the sample is designed specifically to keep non-response low;
- Taking a representative subsample of non-respondents and using it to make inferences about the other non-respondents;
- Using a model to predict values for the non-respondents:
 - Weights implicitly use a model to adjust for unit non-response, and
 - Imputation is used for item non-response; and

- Ignoring the non-response, but it is not recommended [18].

Now the question should be: what kind of response rate will give valid results? The answer depends on the nature of the non-response. For this purpose, let us introduce the following type of non-response [18]:

1. Missing completely at random (MCAR):

The probability of non-response is unrelated to any variable measured on the sample.

2. Non-ignorable:

The probability of non-response depends on the value of a response variable.

If the non-respondents are MCAR, then the respondents are treated as a representative sample of the population and the non-respondents are ignored. If, however, the non-response is non-ignorable, then any results obtained from using only respondents would be worthless.

There are different definitions for calculating the response rate of a survey and each definition gives a very different value. Also, some definitions result in higher response rates than others. One such a definition is, for instance,

$$\frac{\text{number of completed interviews}}{\text{number of units in sample}},$$

under SRS design. Giving guidelines as to what response rate should result in accurate estimates, is careless since there exist surveys with response rates of 95% that have given flawed results [18].

Weighting cells are formed by using categorical variables that are known for all units in the sample and forming subgroups (cells) by cross-classifying the categories of these variables. It is hoped that respondents and non-respondents in the same cell are similar [18]. The weights of the respondents are then adjusted so that the achieved sample represents the intended sample, and hence the population [20].

In the simulation study that will follow later on, a 100% response rate will be assumed ignoring all non-response. Thus, some of the methods for dealing with non-response are only briefly discussed, many more do exist, and will not be considered further in this thesis.

2.4.3 Adjustment of Sample Weights for Non-Coverage and Under-Representation

Non-coverage is the failure of the sampling frame, a list or collection of population units from which the sample is actually selected, to cover the target population. This results in some of the population units having a zero selection probability for the sample [23].

It is important to consider non-coverage and under-representation, especially for surveys conducted in developing countries such as South Africa. In developing countries it is especially challenging to keep an up to date sampling frame and, depending on the parameter of interest to be estimated from the sample, significant differences could result between estimates based on developing-country surveys and those from various other sources. For instance, in this thesis, welfare indices will be estimated. If non-coverage and/or under-representation should occur in the sample and it is not taken into account, the resulting estimates may not give an accurate reflection of the level of poverty and inequality in South Africa.

Under-representation of certain parts of the population (such as young males or small households) appear generally in practice which could lead to biased results if it is ignored. Therefore it should be identified and controlled. Some of the approaches to handling non-coverage and/or under-representation, are:

1. Improved field procedures, and
2. Compensating for non-coverage and/or under-representation through an adjustment of weights.

In this thesis we shall be considering the second approach. The following weight adjustment methods will be considered:

1. Post-Stratification,
2. Cell-Weighting,
3. Calibration Weighting, and
4. Integrated Weighting.

2.4.4 Post-Stratification and Cell-Weighting

Adjustment made by means of post-stratification consists of dividing the sample elements into subgroups called post-strata [20]. After this has been done, an adjustment is made to the weight of each element in a given subgroup. The adjustment is made to reduce variances of estimators involving variables correlated with characteristics used to partition the population into post-strata. If the fixed total for each post-stratum is equal to the expected value of the sample estimate of that total, then the procedure introduces no bias [27]. Post-stratification makes use of a ratio estimator within each subgroup to adjust by the true population count. Let

$$x_{ai} = \begin{cases} 1, & \text{if } i \text{ is a respondent in post-stratum } a \\ 0, & \text{otherwise} \end{cases}.$$

Then let

$$w_i^* = \sum_{a=1}^A w_i x_{ai} \cdot \frac{N_a}{N_{aR}}, \quad (2.8)$$

where A is the number of post-strata, N_a is the population total in post-stratum a , N_{aR} is the population total in post-stratum a based only on respondents and w_i is the design weight of the i th sampling unit. The weight defined in (2.8), w_i^* , is called the post-stratum weight [18].

Cell-weighting and post-stratification work well where population numbers in the interlaced cells are known and the sample is large enough, but population information is often available only at certain levels. It is also effective when cells that are too small or empty appear in the sample. This is where calibration and integrated weighting can be used [20]. Calibration weighting is discussed in section 2.4.4 and integrated weighting is discussed in section 2.4.5.

2.4.5 Calibration Weighting

The calibration technique was introduced by Deville and Särndal in 1992 [7] and by Devill et al. in 1993 [8]. It is a widely used procedure for obtaining improved estimates in sampling surveys by making use of auxiliary information in the form of known population totals to produce a new adjusted set of weights, called calibration weights.

The following notation should be introduced [22]:

- A sample, S , of m households with a total of n persons is drawn from the finite population, U , of M households with a total of N persons. Weighting cells are formed by using categorical variables that are known for all units in the sample and forming subgroups (cells) by cross-classifying the categories of these variables. It is hoped that respondents and non-respondents in the same cell are similar [18]. The weights of the respondents are then adjusted so that the achieved sample represents the intended sample, and hence the population [20]. Let
- m_h , the number of members in household h , $h = 1, \dots, m$.
- n , the number of members sampled, $\sum_{h=1}^m m_h = n$.
- π_k , the inclusion probability of the k th population element.
- $\Pi = \text{diag}(\pi_k)$, the $N \times N$ diagonal matrix of inclusion probabilities.
- $d_k = \frac{1}{\pi_k}$, the design weight of $k \in U$.
- y_k , the study variable.
- $\mathbf{Y} = (y_1, \dots, y_N)'$, the N -vector of values of the study variable.
- x_1, \dots, x_J , the J auxiliary variables.
- $\mathbf{x}_k = (x_{k1}, \dots, x_{kJ})'$, the J -vector for each $k \in U$.
- $\mathbf{X}_T = \sum_{k \in U} \mathbf{x}_k$, the J -vector with known population totals.
- $\hat{\mathbf{X}}_\pi = \sum_{k \in S} d_k \mathbf{x}_k$, the Horvitz-Thompson estimator of the auxiliary variables.

Vectors and matrices for the sample will be denoted by the subscript S .

The auxiliary information can be obtained from external sources such as census data. The calibration estimator is given by

$$\hat{Y}_{cal} = \sum_{k \in S} w_k y_k, \quad (2.9)$$

where w_k are the calibration weights that are as close as possible to the design weights, d_k [22]. The calibration weights are subject to a set of constraints, namely

$$\sum_{k \in S} w_k \mathbf{x}_k = \mathbf{X}_T, \quad (2.10)$$

where the vector \mathbf{X}_T contains the known population totals and \mathbf{x}_k is a vector containing the values of the different auxiliary variables for each element in the population. Equation (2.10) ensures that the sample sum of the weighted auxiliary variables equal the known population total for that variable [20].

Consider a general distance function

$$G(w_k, d_k) = d_k v_k G\left(\frac{w_k}{d_k}\right),$$

that measures the distance between the original weight d_k and the new weight w_k , where v_k is a known positive weight unrelated to d_k [7].

Now, new weights $w_k, k \in S$, have to be found that minimises the average distance for the whole sample,

$$\min_{w_k} \sum_{k \in S} G(w_k, d_k),$$

subject to the constraint in (2.10). From this it follows that the calibration weights are given by

$$w_k = d_k F\left(\frac{\mathbf{x}_k' \lambda_c}{v_k}\right),$$

where $\lambda_c = (\lambda_1, \dots, \lambda_J)'$ is the Lagrange multiplier vector and F is the inverse function of $\frac{dG(\psi)}{d\psi}$ for $\psi = \frac{w_k}{d_k}$ [22]. Thus, the calibration estimator in (2.9) is now given by

$$\hat{Y}_{cal} = \sum_{k \in S} d_k F\left(\frac{\mathbf{x}_k' \lambda_c}{v_k}\right) y_k.$$

Several distance functions have been suggested in the literature, inter alia the linear, exponential (or the so called raking ratio), logit (truncated exponential) and truncated linear methods. In the case of the linear method the calibration weights are given by

$$w_k = d_k \left(1 + \mathbf{x}_k' \lambda_c / v_k \right),$$

where λ_c is determined by the solution to the system

$$\left(\sum_{k \in S} d_k \mathbf{x}_k \mathbf{x}_k' / v_k \right) \lambda_c = \mathbf{X}_T - \hat{\mathbf{X}}_\pi,$$

and v_k is usually set equal to one [20].

The efficiency of the estimator \hat{Y}_{cal} depends on how well the auxiliary variables explain the variability of the variable of interest. Thus, the weights perform well given that there exists a strong correlation between auxiliary variables and study variables [20].

One of the disadvantages of this method is that it may produce weights that are either negative, resulting from an over-constrained system, or large and positive, leading to an increase in the standard error of the estimator. Also, the shortcoming of using a calibration technique for adjusting person weights, is that the weights will usually differ from person to person within the same household. Hence, it does not produce a representative household weight which could be used to estimate household variables of interest. Furthermore, the calibration estimators do not take the household as a cluster into account.

2.4.6 Integrated Weighting

In the past, household surveys generally used separate weighting procedures for estimating person and household characteristics. As a result it produces different sets of weights. Since calibration weighting produces weights that differ between household members, it does not produce a representative household weight. This can introduce some uncertainty in estimating household variables. Integrated linear weighting was developed to achieve a single set of weights that can be used for both person and household estimation [20].

Integrated Weighting: Person Level

Let us assume the finite population U contains M households with a total of N persons. A sample of m households has been drawn with a total of n persons. Let L be an $N \times M$ matrix that links person and household data by [20]

$$L_{kh} = \begin{cases} 1, & k \in h \\ 0, & \text{otherwise} \end{cases}.$$

Here h refers to the household to which person k belongs. A method proposed by Lemaître and Dufour in 1987 [17] replaces X_S with Z_{pp} , where $\{X_S\}_{ij}$ is the (ij) th entry of the $n \times J$ matrix, indicating the value of auxiliary variable j for person i in the sample and $\{Z_{pp}\}_{ij}$ is the proportion of people in the i th chosen household with auxiliary characteristic j . The subscript “pp” denotes the use of person-based auxiliary variables only [22]. The elements of this matrix are given by [20]

$$z_{hj} = \frac{a_{hj}}{m_h},$$

and are defined for person k of household h with m_h members. Note that

$$a_{hj} = \sum_{k \in h} x_{kj},$$

is the total for characteristic j in household h . Thus, the matrix Z_{pp} at person level is defined as

$$Z_{pp} = L_S K_{HS}^{-1} A_{HS},$$

where K_{HS} is a $m \times m$ diagonal matrix containing the household sizes m_h , $h = 1, \dots, m$, and A_{HS} is a $m \times J$ matrix given by

$$A_{HS} = L_S' X_S,$$

that includes the auxiliary variables through the $n \times J$ matrix X_S , aggregated per household [20].

When both person and household auxiliary variables exist, the matrix that already contains the person variable information, can now be extended by adding

columns for each category of the household auxiliary variable under consideration. The entry is then simply the inverse of the household size for the category in which the household falls and zero for all other categories [20]. The Z matrix that includes both person and household auxiliary variables will be denoted by Z_{ph} .

The $n \times 1$ person level vector of weights is

$$\mathbf{W}_S = \Pi_S^{-1} \mathbf{1}_n + \Pi_S^{-1} Z_S \left(Z_S' \Pi_S^{-1} Z_S \right)^{-1} \left(X_T - \hat{X}_\pi \right), \quad (2.11)$$

where

$$\Pi_S = \text{diag}(\pi_k)$$

is a diagonal matrix containing the inclusion probability of the k th sampled element and Z_S denotes Z_{pp} or Z_{ph} . These weights satisfy a set of constraints [20],

$$Z_S' \mathbf{W}_S = \mathbf{X}_T. \quad (2.12)$$

Integrated Weighting: Household Level

The above integrated weights, calculated on person level, can also be calculated on a household-based dataset. The method proposed for the household auxiliary variable case replaces matrices X_S and Z_{pp} by A_{HS} , the matrix of aggregates of the auxiliary characteristics of household members. Furthermore, if reliable population counts are also available for households, this information can be added in the form of additional columns to the matrix A_{HS} such that dummy variables denote whether a household belongs to a certain category or not [20].

Household weights are defined as

$$\mathbf{W}_{HS} = \Pi_{HS}^{-1} \mathbf{1}_m + \Pi_{HS}^{-1} V_{HS}^{-1} A_{HS} \left(A_{HS}' \Pi_{HS}^{-1} V_{HS}^{-1} A_{HS} \right)^{-1} \left(X_T - \hat{X}_\pi \right), \quad (2.13)$$

and are subjected to the same set of constraints (2.12) as the person weights.

Now all members of a household retain the same weight and when the weights are multiplied by the number of persons in each category of a person-level auxiliary variable, the weighting estimates agree with the marginal population totals of that variable at person level [20].

Finally, the link between the person-based weights in (2.11) and the household-based weights in (2.13) is given by either

$$\mathbf{W}_{HS} = K_{HS}^{-1} L_S' \mathbf{W}_S,$$

or

$$\mathbf{W}_S = L_S K_{HS} \mathbf{W}_{HS},$$

where K_{HS} is a $m \times m$ diagonal matrix containing the household sizes. It has been shown that the integrated weighting technique based on person level data yields the same final weights than the technique based on household level data. Thus, the decision of which data to use relies on the current situation, the auxiliary information available as well as the desired estimators.

2.5 Estimation

Consider a two-stage complex design with

- h , the stratum index,
- j , the psu index, and
- i , the ultimate sampling unit (usu) index.

Let H be the number of strata. Suppose that n_h psu's are drawn from the h th stratum and m_{hj} usu's are selected in the j th psu of the h th stratum. Consider two variables, x and y , with values

- x_{hji} and y_{hji} for the (h, j, i) th record, and
- w_{hji} the design weight associated with the (h, j, i) th record.

Note that

$$\sum_h \sum_j \sum_i w_{hji} = N,$$

is the number of elements (usu's) in the population [21]. The population totals of the two variables are estimated by

$$\hat{t}_y = \sum_h \sum_j \hat{t}_{yhj} = \sum_h \sum_j \sum_i w_{hji} y_{hji}, \quad (2.14)$$

and

$$\hat{t}_x = \sum_h \sum_j \hat{t}_{xhj} = \sum_h \sum_j \sum_i w_{hji} x_{hji}. \quad (2.15)$$

Since some of the welfare indices are estimated by means of a ratio or proportion it is necessary to define the ratio estimator and proportion estimator in terms of the design weights as well. The ratio estimator is given by

$$\hat{B} = \frac{\sum_h \sum_j \sum_i w_{hji} y_{hji}}{\sum_h \sum_j \sum_i w_{hji} x_{hji}}, \quad (2.16)$$

which can be simplified to

$$\hat{B} = \frac{\hat{t}_y}{\hat{t}_x},$$

from equations (2.14) and (2.15) [21]. Suppose the proportion of grade 12 learners with marks for mathematics above 60% that intend to enroll at university has to be estimated. Let x_{hji} and y_{hji} be indicator variables where x_{hji} denotes the (h, j, i) th learner's (usu's) intent to enroll at university and y_{hji} denotes whether the (h, j, i) th learner (usu) intends to enroll at university and has a mark above 60% in mathematics. Then

$$\hat{p} = \frac{\sum_h \sum_j \sum_i w_{hji} y_{hji}}{\sum_h \sum_j \sum_i w_{hji} x_{hji}}. \quad (2.17)$$

Note that (2.17) is the same as (2.16). The usefulness of design weights comes into its own in the above equations [21].

2.6 Conclusion

The chapter commenced with a brief overview of probability sampling methods, namely simple random sampling, systematic sampling, stratified random sampling and cluster sampling. This was followed by a discussion on complex sampling. Let us recall that a complex sample is defined as a stratified multistage cluster sample

and potential hazards one likely will stumble upon are weighting, non-response and errors that occur in estimation from a complex sample design.

The next section was devoted to weighting in general and in a complex sampling design. It is a necessary method for dealing with the effects of stratification and clustering on estimates. Recall that the inclusion probability of the i th observation unit of the j th psu in the h th stratum is denoted by p_{hji} . Thus its design weight is simply given by

$$w_{hji} = \frac{1}{p_{hji}}, \quad h = 1, \dots, H, j = 1, \dots, n_h, i = 1, \dots, n_{hj},$$

since the design weight is defined as the reciprocal of the inclusion probability. Weights are developed in different stages to ensure that the sample represents the population as closely as possible. Shortly, the first stage sees the development of the design weight, the second stage adjusts the design weights for non-response and the weighting is concluded with the adjustment of the weighted estimates to match known population totals through the use of auxiliary information available for certain key variables.

The adjustment for non-response can be done by the method of weighting cells. Since it is accepted that for the purpose of this thesis there is no non-response in the data the method was only discussed briefly. The adjustment in the final stage of weighting can be done through various methods of which only calibration weighting and integrated weighting were discussed in detail. Another method mentioned and briefly described was post-stratification. Calibration weighting makes use of auxiliary information in the form of known population totals to produce a new adjusted set of weights referred to as the calibration weights. This auxiliary information can be obtained from external sources such as census data. One of the shortcomings of calibration weighting is that it does not produce a representative household weight. Hence, integrated weighting was developed to achieve a single set of weights that can be used for both person and household estimation.

Thus, if $w_{hji,b}$ represented the design weight, $w_{hji,nr}$ represented the weight compensating for non-response and $w_{hji,nc}$ represented the weight compensating for non-coverage, then the overall weight of an observation unit will be calculated as

$$w_{hji} = w_{hji,b} \cdot w_{hji,nr} \cdot w_{hji,nc}.$$

Once the weights have been calculated they are applied directly in the calculation of estimators. Given that the weights have been adjusted correctly, they should sum to the total number of elements in the population. This forms the basis for the calculation of any estimators.

Chapter 3

Resampling Techniques

Estimating statistics such as population means and totals from complex surveys is easily done through the use of the design weights, but estimating the associated variances can be difficult [18]. In a complex survey with various levels of stratification and clustering, the variances are calculated at each level and then combined towards the end. Variance estimation formulas have been derived for many statistics for a selection of sampling designs. Some formula, as in simple random samples, are quite simple while other formulas, as in multistage cluster samples without replacement for instance, tend to be quite intricate. There are certain quantities for which no variance formula has been derived at all, for example the median, in other sampling designs than simple random samples [18].

A well known and commonly used method for estimating variances is Taylor linearization. It is applicable to general sampling designs which permit unbiased variance estimation for linear estimators. It is also computationally simpler than resampling methods such as the jackknife discussed below [6]. Many of the nonlinear parameters θ can be expressed as smooth functions of population means or totals of suitably defined variables [15].

Resampling methods are widely used techniques in the independent and identically distributed (i.i.d.) case [26]. They are, however, not limited to the this situation [4]. They provide a simple, robust approach to the estimation of sampling variances as well as the conducting of significance tests for complex survey data. Let the population parameter of interest be denoted by θ . Since the analyst usually does not know the entire population, θ is estimated from a sample taken from the population. The sample is denoted by $\mathbf{y} = (y_1, \dots, y_n)$ and the estimator is denoted

by $\hat{\theta} = \hat{\theta}(y_1, \dots, y_n)$. For example, let the variable of interest be income. Then \mathbf{y} is the sample that contains the measures taken on the variable of interest, y_i is the income of person i , and $\hat{\theta}$ might be the average income or the Gini coefficient. In a nutshell, these methods consist of estimating the variance of the estimators of the population parameters that are of interest to the analyst. This is done by taking subsamples from the sample \mathbf{y} and calculating $\hat{\theta}$ for each subsample. The variability between the calculated estimators is then used to estimate the variance of the initial estimator [27].

Two methods will be considered:

1. The Jackknife, and
2. The Bootstrap.

In both methods we have a sample $\mathbf{y} = (y_1, y_2, \dots, y_n)$ of size n and an estimator $\hat{\theta} = s(\mathbf{y})$ where $s(\cdot)$ is some statistic measured on the sample \mathbf{y} . In the sections referring to application of the resampling methods in complex surveys, there are H strata, $h = 1, \dots, H$, with j psu's in stratum h , $j = 1, \dots, n_h$. An observation i in the j th psu in the h th stratum will be denoted by y_{hji} .

3.1 The Jackknife

The jackknife method predates the bootstrap method in the estimation of bias and standard errors of an estimator $\hat{\theta}$. Its name was used by Tukey in 1958 as a way of conveying the broad usefulness of this technique [13]. A jackknife is synonymous to a penknife or a switchblade, which is a multipurpose knife that can perform the functions of a number of more specialized knives. Thus, the jackknife can be used as a substitute for a variety of more specialized techniques. In general, the method proceeds by deleting the i th sample observation and calculating a replicate, $\hat{\theta}_{(i)}$, from the $(1, \dots, i-1, i+1, \dots, n)$ observations in the i th jackknife sample, $i = 1, \dots, n$ [10]. This technique has at most n repetitions and as such it is much less computationally intensive than the bootstrap.

3.1.1 Estimation of Bias

Simple Random Sampling

The bias of an estimator, $\hat{\theta}$, is defined as the difference between the expected value of $\hat{\theta}$ and the parameter being estimated, θ [10]. The jackknife estimate of bias was developed by Quenouille in 1949 [13]. Let $\hat{\theta}$ be an estimator of some parameter of interest, θ . The estimator is based on a sample of independent and identically distributed (i.i.d.) random variables, $\hat{\theta} = \hat{\theta}(y_1, \dots, y_n)$. Suppose that

$$E_{\theta}(\hat{\theta}) = \theta + \text{bias}_{\theta}(\hat{\theta}), \quad (3.1)$$

where (3.1) is the expected value of $\hat{\theta}$ as an estimator of θ and $\text{bias}_{\theta}(\hat{\theta})$ is the bias of $\hat{\theta}$ [13]. In an ideal situation we would want this expected value of the estimator to be exactly equal to the parameter of interest, θ . Then we would have an unbiased estimator. Thus, if $\hat{\theta}$ is a good estimator of θ , we would expect $\text{bias}_{\theta}(\hat{\theta})$ to be close to zero. Let $\hat{\theta}_{(i)}$ be the estimator of θ calculated from the sample after omitting the i th observation:

$$\hat{\theta}_{(i)} = \hat{\theta}(y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n). \quad (3.2)$$

The jackknife estimator of the bias of $\hat{\theta}$ is then

$$\widehat{\text{bias}}_{JK}(\hat{\theta}) = (n-1)(\tilde{\theta}_{JK} - \hat{\theta}), \quad (3.3)$$

where $\tilde{\theta}_{JK}$ is the average of the jackknife replicates, $\{\hat{\theta}_{(i)}\}$, $i = 1, \dots, n$, and $\hat{\theta}$ is the estimator of θ [13].

The motivation behind (3.3) is the assumption that (3.1) can be expressed as a series involving powers of $\frac{1}{n}$ [13]. Let us first assume that for any n

$$E_{\theta}(\hat{\theta}) = \theta + \frac{a_1(\theta)}{n}, \quad (3.4)$$

where $a_1(\theta)$ can depend on θ or on the distribution of $\{y_i\}$, but not on the size of the sample, n . In this case the bias, $\text{bias}_{\theta}(\hat{\theta})$, is equal to $\frac{a_1(\theta)}{n}$. Then,

$$E_{\theta}(\tilde{\theta}_{JK}) = \frac{1}{n} \sum_{i=1}^n E_{\theta}(\hat{\theta}_{(i)}). \quad (3.5)$$

Recall that $\tilde{\theta}_{JK}$ is the average of the jackknife replicates, $\{\hat{\theta}_{(i)}\}$. Now,

$$E_{\theta}(\tilde{\theta}_{JK}) = \theta + \frac{a_1(\theta)}{n-1},$$

because the replicates are based on $(n-1)$ observations. The formula in (3.3) then follows directly from

$$(n-1) E_{\theta}(\tilde{\theta}_{JK} - \hat{\theta}) = (n-1) \left[\frac{a_1(\theta)}{(n-1)} - \frac{a_1(\theta)}{n} \right] = \frac{a_1(\theta)}{n}. \quad (3.6)$$

Formula (3.6) shows that the jackknife produces an unbiased estimator of the bias of $\hat{\theta}$ in the simple case illustrated above, since

$$bias_{\theta}(\hat{\theta}) = E_{\theta}(\hat{\theta}) - \theta = \frac{a_1(\theta)}{n},$$

from the definition of bias and from (3.4) [13]. However, in the more general case we find that the jackknife does not produce an unbiased estimator of bias. In general, let

$$E_{\theta}(\hat{\theta}) = \theta + \frac{a_1(\theta)}{n} + \frac{a_2(\theta)}{n^2} + \frac{a_3(\theta)}{n^3} + \dots,$$

or

$$bias_{\theta}(\hat{\theta}) = \frac{a_1(\theta)}{n} + \frac{a_2(\theta)}{n^2} + \frac{a_3(\theta)}{n^3} + \dots, \quad (3.7)$$

where $a_1(\theta)$, $a_2(\theta)$, $a_3(\theta)$, \dots can depend on θ or on the distribution of $\{y_i\}$, but not on the size of the sample. Exactly as shown in the steps above, we arrive at the expected value of the jackknife estimator of bias [13]:

$$E_{\theta}(\widehat{bias}_{JK}(\hat{\theta})) = \frac{a_1(\theta)}{n} + \frac{(2n-1)a_2(\theta)}{n^2(n-1)} + \frac{(3n^2-3n+1)a_3(\theta)}{n^3(n-1)^2} + \dots. \quad (3.8)$$

Notice the difference between (3.6) and (3.8). It is clear from (3.8) that the jackknife estimator of bias is not in general an unbiased estimator. This brings us to a bias-corrected version of the estimator, $\hat{\theta}$. Define

$$\hat{\theta}_{jack} = \hat{\theta} - \widehat{bias}_{JK}(\hat{\theta}) = n\hat{\theta} - (n-1)\tilde{\theta}_{JK}. \quad (3.9)$$

Then it follows that

$$E_{\theta}(\hat{\theta}_{jack}) \approx \theta - \frac{a_2(\theta)}{n^2} - \frac{2a_3(\theta)}{n^3} - \dots, \quad (3.10)$$

as n grows large [13]. Since the denominators in (3.10) go to zero faster than $\frac{1}{n}$ as n grows large, it follows that the bias-corrected estimator has smaller bias than $\hat{\theta}$ if n is sufficiently large.

Complex Sampling

Since the jackknife method will only be used for the estimation of the standard error of bootstrap replicates in this study, the jackknife estimation of bias is only discussed for a simple random sample.

3.1.2 Estimation of Standard Error

Simple Random Sampling

Let us firstly consider the application in a simple random sample of size n . If we want to estimate the mean of the population, μ , we would make use of $\hat{\theta}_{(i)} = \bar{y}_{(i)}$, the mean of the sample where the i th observation has been deleted. The jackknife estimator of the variance of $\hat{\theta}$ is

$$\hat{V}_{JK}(\hat{\theta}) = \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta})^2. \quad (3.11)$$

This is referred to as the delete-1-jackknife estimator due to only one observation at a time being deleted. Other forms of the jackknife do exist, called the delete-d-jackknife, but will not be considered here. Note that this method of resampling requires at most n steps [18].

In the derivation of (3.11), Tukey assumed that

$$\hat{\theta} \approx \frac{1}{n} \sum_{i=1}^n \phi(y_i),$$

where $\phi(\cdot)$ is a function of the data. This suggests that

$$V(\hat{\theta}) \approx \frac{V(\phi(y_1))}{n},$$

since the $\{y_i\}$ are i.i.d. We do not generally know what $\phi(\cdot)$ is, but it is possible to find substitutes for $\phi(y_i)$ [13]. Define pseudo-values

$$\Phi_i = \hat{\theta} + (n-1)(\hat{\theta} - \hat{\theta}_{(i)}),$$

where $\hat{\theta}_{(i)}$ is the jackknife replicate calculated after the i th sample value was deleted. These n pseudo-values perform the same role as the n $\phi(y_i)$ values. The sample variance of these pseudo-values can be used to estimate the variance of $\phi(y_1)$ which can then be used to estimate the variance of $\hat{\theta}$ [13]. Note that

$$\frac{1}{n} \sum_{i=1}^n \Phi_i = n\hat{\theta} + \frac{n-1}{n} \sum_{i=1}^n \hat{\theta}_{(i)} = \hat{\theta}_{jack},$$

which shows that the average of the pseudo-values are simply equal to the bias-corrected version of $\hat{\theta}$. Now, the sample variance of the pseudo-values is given by

$$\frac{1}{n-1} \sum_{i=1}^n (\Phi_i - \bar{\Phi})^2 = (n-1) \sum_{i=1}^n (\hat{\theta}_{(i)} - \hat{\theta}(\cdot))^2. \quad (3.12)$$

From (3.12) it is easy to get the jackknife estimator of variance [13].

Complex Sampling

Now consider the application of jackknife in a stratified multistage cluster sample (complex sample) with H strata and n_h psu's (primary sampling units) in each stratum. Let $\hat{\theta}_{(hj)}$ be the estimator of θ , obtained from the sample after the data from the j th psu in the h th stratum has been deleted, $j = 1, \dots, n_h$ and $h = 1, \dots, H$, and the weights of all other units from the h th stratum have been inflated by a factor of $\frac{n_h}{n_h-1}$ [27]. Thus

$$w_{i(hj)} = \begin{cases} w_i, & i \notin h \\ w_i \cdot \frac{n_h}{(n_h-1)}, & i \in h, i \neq j \\ 0, & i \in (h, j) \end{cases} \quad (3.13)$$

These new weights are then used in the calculation of the estimator. An estimate of the total would typically be calculated as

$$\hat{y}_{(hj)} = \sum_{h=1}^H \sum_{j=1}^{n_h} \sum_{i \in j} w_{i(hj)} y_{hji}, \quad (3.14)$$

for each $j \in h$ and $h = 1, \dots, H$. The jackknife variance estimator of the sampling variance of \hat{y} is given by

$$V_{JK}(\hat{y}) = \sum_{h=1}^H \left(\frac{n_h - 1}{n_h} \right) \sum_{j=1}^{n_h} (\hat{y}_{(hj)} - \hat{y})^2. \quad (3.15)$$

where \hat{y} is the estimated population total. One of the advantages of using the jackknife variance estimator is the application to more complex estimators [27]. For a general population parameter, θ , the jackknife variance estimator of the sampling variance of $\hat{\theta}$ is given by

$$\hat{V}_{JK}(\hat{\theta}) = \sum_{h=1}^H \left(\frac{n_h - 1}{n_h} \right) \sum_{j=1}^{n_h} (\hat{\theta}_{(hj)} - \hat{\theta})^2. \quad (3.16)$$

Sometimes the formula subtracts the mean of the jackknife replicates, $\tilde{\theta}_{JK}$, rather than calculating $\hat{\theta}$ from the original sample in which case the jackknife variance estimator would be

$$\tilde{V}_{JK}(\hat{\theta}) = \sum_{h=1}^H \left(\frac{n_h - 1}{n_h} \right) \sum_{j=1}^{n_h} (\hat{\theta}_{(hj)} - \tilde{\theta}_{JK})^2, \quad (3.17)$$

where $\tilde{\theta}_{JK}$ is the mean of the jackknife replicates, $\hat{\theta}_{(hj)}$. $\hat{V}_{JK}(\hat{\theta})$ and $\tilde{V}_{JK}(\hat{\theta})$, in both simple random sampling and complex sampling, shows close resemblance to the formula for calculating the sample variance,

$$\hat{V}(y) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (3.18)$$

The difference between (3.18) and (3.11) and (3.17) lies in the factors $\frac{n-1}{n}$ (3.11) and $\frac{n_h-1}{n_h}$ (3.17) instead of $\frac{1}{n-1}$ or $\frac{1}{n}$ (3.18). The former factors are much larger than the latter factors used in the sample variance and are termed “inflation factors”. These “inflation factors” are needed, because the jackknife deviations $(\hat{\theta}_{(i)} - \hat{\theta})^2$,

and $\left(\hat{\theta}_{(hj)} - \hat{\theta}\right)^2$ in complex samples, are much smaller than the deviations for other resampling techniques such as the bootstrap, since the jackknife sample is more similar to the original sample than a typical bootstrap sample. These “inflation factors” are also needed to compare the jackknife variance estimator to other variance estimators such as the bootstrap variance estimator [10].

Consider the case where R jackknife populations are formed from the original sample. Let the r th jackknife population be denoted by

$$Y^r = \{Y_{r1}, \dots, Y_{rn}\},$$

and let the estimator of the parameter of interest for the r th jackknife population be defined as

$$\hat{\theta}^r = \hat{\theta}(Y^r).$$

The jackknife technique will be applied in each of the R jackknife populations. Consider the r th jackknife population and let $\hat{\theta}_{(hj)}^r$ be the jackknife estimator obtained after the j th psu of the h th stratum in the r th population has been deleted. The design weights of the units will be adjusted as follows:

$$w_{i(hj)}^r = \begin{cases} w_i^r, & i \notin h \\ w_i^r \cdot \frac{n_h}{(n_h-1)}, & i \in h, i \notin j \\ 0, & i \in (h, j) \end{cases},$$

and these weights will be used in the calculation of $\hat{\theta}_{(hj)}^r$. The jackknife variance estimator of $\hat{\theta}^r$ is given by

$$\hat{V}_{JK}(\hat{\theta}^r) = \sum_{h=1}^H \left(\frac{n_h - 1}{n_h} \right) \sum_{j=1}^{n_h} \left(\hat{\theta}_{(hj)}^r - \hat{\theta}^r \right)^2.$$

The other form of the variance estimator that can be used, is

$$\tilde{V}_{JK}(\hat{\theta}^r) = \sum_{h=1}^H \left(\frac{n_h - 1}{n_h} \right) \sum_{j=1}^{n_h} \left(\hat{\theta}_{(hj)}^r - \tilde{\theta}^r \right)^2,$$

where

$$\tilde{\theta}^r = \frac{1}{n_h} \sum_{j=1}^{n_h} \hat{\theta}_{(hj)}^r.$$

The application of the jackknife technique in each jackknife population will result in R jackknife variance estimates,

$$\hat{V}_{JK}(\hat{\theta}^1), \dots, \hat{V}_{JK}(\hat{\theta}^R).$$

The overall jackknife variance estimator of $\hat{\theta}$ is then simply the average of these variance estimates,

$$\hat{V}_{JK}(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R \hat{V}_{JK}(\hat{\theta}^r).$$

Advantages of the jackknife that have to be mentioned are that the same procedure is used to estimate the variance of every statistic for which the jackknife can be used and it provides a consistent estimator of the variance when θ is a smooth function of population totals. On the other hand, it performs miserably if the statistic is not smooth [18]. Results obtained when applied in unequal probability sampling designs where sampling is done without replacement should not be trusted since little is known about the performance of the jackknife method under these circumstances [18].

3.1.3 A comparison of the Jackknife method and Linearization Method

The comparison is made for the sample mean, \bar{y} , of the i.i.d. random variables, y_i . The estimator for variance from the linearization method is given by

$$\hat{V}_L(g(\bar{y})) = \left[g'(\bar{y}) \right]^2 \frac{1}{n(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2, \quad (3.19)$$

where $g(\bar{y}) = \hat{\theta}$. The jackknife estimator of variance is given by

$$\hat{V}_{JK}(g(\bar{y})) = \frac{n-1}{n} \sum_{i=1}^n (g(\bar{y}_{(i)}) - g(\bar{y})), \quad (3.20)$$

where $g(\bar{y}_{(i)}) = \hat{\theta}_{(i)}$ and $g(\bar{y})$ is the mean of the $\{g(\bar{y}_{(i)})\}$. Note that

$$\bar{y}_{(i)} = \frac{1}{n-1} \sum_{j \neq i}^n y_j = \bar{y} - \frac{1}{n-1} (y_i - \bar{y}). \quad (3.21)$$

If we perform a Taylor series expansion on (3.21) we get

$$g(\bar{y}_{(i)}) \approx g(\bar{y}) + (\bar{y}_{(i)} - \bar{y}) g'(\bar{y}), \quad (3.22)$$

and hence

$$g(\bar{y}) = \frac{1}{n} \sum_{i=1}^n g(\bar{y}_{(i)}) \approx g(\bar{y}). \quad (3.23)$$

If we substitute the approximations in (3.22) and (3.23) into (3.19) we get

$$\hat{V}_{JK}(g(\bar{y})) \approx [g'(\bar{y})]^2 \frac{1}{n(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2,$$

which is equal to the linearization estimator in (3.18). Thus, when $\hat{\theta}$ is the sample mean, the jackknife method and the linearization method are approximately equal [13].

3.2 The Bootstrap

The bootstrap resampling method was introduced in 1979 as a computer intensive method for estimating the variance of an estimator, $\hat{\theta}$. A pleasing property of this resampling method is that there is no need to derive theoretical variances and the bootstrap estimate is available regardless of how mathematically complicated the estimator may be [10].

In this study the survey data will be used as a surrogate population from which R samples will be taken. These samples are called bootstrap populations. The bootstrap procedure will be applied to each bootstrap population and then the results will be aggregated over the R bootstrap populations. This procedure is

described in more detail in sections 3.2.1 and 3.2.2. Note that in the notation provided there is implied that the same number of bootstrap resamples, B , are used for each bootstrap population although one could use a different number of bootstrap samples for each bootstrap population, namely B_r . This is omitted in order to keep the notation simple.

3.2.1 Estimation of Bias

Simple Random Sampling

The bias of $\hat{\theta} = \hat{\theta}(y_1, \dots, y_n)$ as an estimator of θ is given by

$$bias_F(\hat{\theta}) = E_F(\hat{\theta}) - \theta, \quad (3.24)$$

where subscript F in (3.24) denotes the probability distribution from which the sample, \mathbf{y} , was taken. The aim is to have a small bias. Plug-in estimates, as $\hat{\theta}$ usually is, are not necessarily unbiased, but their biases tend to be small in comparison to their standard errors which is one of the pleasing properties of plug-in estimates [10].

The bootstrap can be used to assess the bias of an estimator, $\hat{\theta}$, and is defined by making use of the plug-in principle and replacing F in (3.24) with \hat{F} ,

$$bias_{\hat{F}}(\hat{\theta}) = E_{\hat{F}}(\hat{\theta}) - \hat{\theta}, \quad (3.25)$$

where \hat{F} in (3.25) denotes the empirical distribution function. The bootstrap method starts by generating B independent bootstrap samples, $\mathbf{y}_1^*, \dots, \mathbf{y}_B^*$, where B is a large number. For each bootstrap sample the bootstrap replicate, $\hat{\theta}_b^* = \hat{\theta}(\mathbf{y}_b^*)$, $b = 1, \dots, B$, is calculated. The bootstrap approximation of $E_{\hat{F}}(\hat{\theta})$ is given by

$$\tilde{\theta}^* = \frac{1}{B} \sum_{b_r=1}^B \hat{\theta}_{b_r}^*,$$

the average of the B bootstrap replicates [10]. The bootstrap estimate of bias is then

$$\widehat{bias}_B(\hat{\theta}) = \tilde{\theta}^* - \hat{\theta}. \quad (3.26)$$

Complex Sampling

Consider a stratified multistage cluster sample with H strata and n_h psu's in stratum h . The bootstrap method as described for a simple random sample is applied independently in each stratum by selecting m_h psu's with replacement and defining the bootstrap weight

$$w_{hji}^* = w_{hji} \left[\left\{ 1 - \sqrt{\left(\frac{m_h}{n_h - 1} \right)} \right\} + \left(\sqrt{\frac{m_h}{n_h - 1}} \right) \left(\frac{n_h}{m_h} \right) m_{hj}^* \right],$$

where w_{hji} is the original design weight of the i th unit in the j th psu of the h th stratum and m_{hj}^* is the number of times the j th psu occurred in the bootstrap sample. The bootstrap weights are then used to calculate the bootstrap replicates, $\{\hat{\theta}_b^*\}$. Let R be the number of bootstrap populations simulated from a surrogate population. The B bootstrap replicates of the estimator of the parameter of interest are given by

$$\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$$

and the estimated bias of the r th bootstrap population is given by

$$\widehat{bias}_B(\hat{\theta}_r^*) = \tilde{\theta}_r^* - \hat{\theta}_r^*,$$

where

$$\tilde{\theta}_r^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_{br}^*.$$

This results in R bias estimates

$$\widehat{bias}_B(\hat{\theta}_1^*), \widehat{bias}_B(\hat{\theta}_2^*), \dots, \widehat{bias}_B(\hat{\theta}_R^*),$$

and the overall estimate of bias is calculated as the average of the R estimates

$$\widehat{bias}_B(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R \widehat{bias}_B(\hat{\theta}_r^*).$$

3.2.2 Estimation of Standard Error

Simple Random Sampling

Firstly consider the application of the bootstrap in a simple random sample of size n . The sample, \mathbf{y} , is then treated as if it were a population and we resample from the sample a large number of times, say B , with replacement. At each resample a bootstrap sample is formed, $\mathbf{y}_b^* = (y_1^*, \dots, y_n^*)$, $b = 1, \dots, B$. If the sample is similar to the underlying population then the bootstrap samples generated from it should reproduce properties similar to samples drawn directly from the original population [18]. For each of the B bootstrap samples a bootstrap replicate, $\hat{\theta}_b^* = \hat{\theta}(y_1^*, \dots, y_n^*)$, $b = 1, \dots, B$, is calculated. The bootstrap estimator of the variance of $\hat{\theta}$ is then simply the variance of the B replicates

$$\hat{V}_B(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_b^* - \hat{\theta} \right)^2. \quad (3.27)$$

It should be mentioned that some literature use the fraction $\frac{1}{B}$ instead of $\frac{1}{B-1}$. Since B is typically a large number it follows that $B-1 \approx B$. Another difference of opinion existing in the literature is the value subtracted from the bootstrap replicates. In (3.27) $\hat{\theta}$ is the estimator calculated on the original sample in contrast to the alternative use of the mean of the bootstrap replicates

$$\tilde{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*, \quad (3.28)$$

in which case the bootstrap estimate of variance would be

$$\tilde{V}_B(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_b^* - \tilde{\theta}^* \right)^2. \quad (3.29)$$

Complex Sampling

The extension of the bootstrap method to stratified multistage sampling proceeds by applying the bootstrap independently in each stratum. Let H be the number of strata and let each stratum contain n_h psu's. In each stratum draw a simple random sample of m_h psu's with replacement and let m_{hj}^* be the number of times the (hj) -th sample psu is selected [26]. The bootstrap weights are defined as

$$w_{hji}^* = w_{hji} \left[\left\{ 1 - \sqrt{\left(\frac{m_h}{n_h - 1} \right)} \right\} + \left(\sqrt{\frac{m_h}{n_h - 1}} \right) \left(\frac{n_h}{m_h} \right) m_{hj}^* \right], \quad (3.30)$$

where w_{hji} is the sampling weight of the i th observation in the j th psu of the h th stratum. Calculate $\hat{\theta}^*$, the bootstrap estimator of $\hat{\theta}$ using these bootstrap weights. Replicate this procedure independently a large number of times, B , and calculate the corresponding bootstrap estimates $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$. Finally, the bootstrap variance estimator is approximated by [26]

$$\hat{V}_B(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^B \left(\hat{\theta}_b^* - \hat{\theta} \right)^2. \quad (3.31)$$

An important point to make is that the bootstrap variance estimator is unbiased when $\hat{\theta}$ is a linear function [27].

Consider the case where R bootstrap populations are formed from the surrogate population. Let the r th bootstrap population be denoted by

$$Y_r^* = \{Y_{r1}^*, \dots, Y_{rn}^*\},$$

and let the parameter of interest for the r th bootstrap population be defined as

$$\hat{\theta}_r^* = \hat{\theta}(Y_r^*).$$

In each of the R bootstrap populations the bootstrap technique will be applied independently in each of the H strata. Assume that $m_h = n_h - 1$ psu's will be selected with replacement in each stratum. Let us consider the r th bootstrap population and let $m_{hjb_r}^*$ be the number of times the j th psu of the h th stratum was sampled for the b th bootstrap sample. The new design weights of the units will be calculated as

$$w_{hji b_r}^* = w_{hji} \cdot \left[\left(\frac{n_h}{n_h - 1} \right) \cdot m_{hjb_r}^* \right], \quad b_r = 1, \dots, B,$$

where w_{hji} is the design weight of the i th unit of the j th psu in the h th stratum. This will be repeated a large number of times, B , resulting in B bootstrap estimators $\{\hat{\theta}_{b_r}^*\}$. The bootstrap estimated variance of $\hat{\theta}_r$ is given by

$$\tilde{V}_B \left(\hat{\theta}_r^* \right) = \frac{1}{B-1} \sum_{b_r=1}^B \left(\hat{\theta}_{b_r}^* - \tilde{\theta}_r^* \right)^2,$$

where

$$\tilde{\theta}_r^* = \frac{1}{B} \sum_{b_r=1}^B \hat{\theta}_{b_r}^*,$$

or by

$$\hat{V}_B \left(\hat{\theta}_r^* \right) = \frac{1}{B-1} \sum_{b_r=1}^B \left(\hat{\theta}_{b_r}^* - \hat{\theta}_r^* \right)^2.$$

After the bootstrap technique has been applied in each bootstrap population, the result will be R bootstrap estimated variances,

$$\hat{V}_B \left(\hat{\theta}_1^* \right), \dots, \hat{V}_B \left(\hat{\theta}_R^* \right).$$

The overall bootstrap variance estimator is simply the average of these variance estimates,

$$\hat{V}_B \left(\hat{\theta} \right) = \frac{1}{R} \sum_{r=1}^R \hat{V}_B \left(\hat{\theta}_r^* \right).$$

The bootstrap estimation method has many advantages. One of them is its applicability to non smooth functions in general sampling designs. It works well in finding confidence intervals directly, but this will be discussed at a later stage. A disadvantage is certainly the number of computations required since B is usually very large [18], but given the computing power currently available, this has become less of a problem.

3.2.3 The choice of B

Until now it has only been stated that B is typically a large number. How large is large enough though? This question does not have a trivial answer. The precision

of the variance estimator increases as the size of B increases, but the bootstrap is computer intensive and the resources needed to carry out the resampling also increase as B increases [27].

There are two conflicting arguments on the size of B . The first argument is from Efron and Tibshirani [10] and is based on the unconditional coefficient of variation of $\hat{V}_B(\hat{\theta})$. Based on this argument only a small number of resamples, B as small as 25 and hardly ever greater than 200, is required to achieve reasonable results [10]. The second argument is by Booth and Sarkar [3], basing their argument on a conditional coefficient of variation. They use only resampling variability in their argument stating that Monte Carlo error (the error due to taking bootstrap samples) should not be allowed to determine the conclusions of an analysis. This argument arrived at the conclusion that many more bootstrap samples are needed for reasonable results, even B as large as 800 [3].

In our application of the bootstrap resampling technique we shall be following the argument by Booth and Sarkar and make use of a B much larger than the size proposed by Efron.

3.2.4 The choice of m_h

After considering extensively the choice of B we have yet to determine the choice of the bootstrap sample sizes in complex surveys. Let us recall the formula for the bootstrap weights (3.27),

$$w_{hji}^* = w_{hji} \left[\left\{ 1 - \sqrt{\left(\frac{m_h}{n_h - 1} \right)} \right\} + \left(\sqrt{\frac{m_h}{n_h - 1}} \right) \left(\frac{n_h}{m_h} \right) m_{hj}^* \right]. \quad (3.32)$$

When $m_h = n_h - 1$ [27], the formula reduces to

$$w_{hji}^* = w_{hji} \left[\left(\frac{n_h}{n_h - 1} \right) m_{hj}^* \right]. \quad (3.33)$$

For this choice of m_h , replicate b , $b = 1, \dots, B$, can be seen as consisting of m_{hj}^* copies of each unit in psu j in stratum h where $j = 1, \dots, n_h$ and $h = 1, \dots, H$. Even in this simplified formula the factor $\frac{n_h}{n_h - 1}$ is still used. Although it can be ignored in many bootstrap applications, it has to be included when n_h is small. This is often the case in complex surveys and the factor helps to avoid the introduction of bias

into the variance estimator [27].

3.2.5 Conclusion

In the case of smooth functions the jackknife variance estimator performs as well as the linearization estimator. The bootstrap estimator tends to overestimate the variance while the jackknife and linearization estimators seem to be more accurate and stable [15]. The jackknife estimator has been proven to be inconsistent for nonsmooth functions. The bootstrap is applicable to general designs which makes it an easy and convenient method to use [15].

The jackknife method is much less computationally intensive since it requires at most a number of resamples as large as the size of the sample on which it is applied. The bootstrap method requires many more resamples, but it is the method that is being used at present and has, in a sense, replaced the jackknife.

Chapter 4

Confidence Intervals

Calculating confidence intervals is an important part of data analysis. Their value as an analysis tool stems from the fact that they combine point estimation and hypothesis testing into a single inferential statement of great intuitive appeal [9]. Until now we have discussed the estimation of standard errors through various resampling techniques. These standard errors are often used to form approximate confidence intervals for a parameter of interest, say θ .

In this chapter we shall describe different techniques for the construction of confidence intervals using the bootstrap. Firstly we shall give a brief overview of the standard asymptotic interval followed by discussions on the percentile interval as well as the bootstrap- t interval. The chapter will be concluded with a short summary of the advantages and disadvantages of each technique as well as remarks on which techniques we shall be using in our application.

4.1 The Standard (Asymptotic) Interval

Let the parameter of interest, θ , be estimated by $\hat{\theta}$ and suppose $\hat{\theta}$ is approximately normally distributed with expected value θ and estimated standard error \widehat{se} . An approximate $100(1 - 2\alpha)\%$ confidence interval for θ is then given by

$$\left[\hat{\theta} - z^{(1-\alpha)} \widehat{se}; \hat{\theta} - z^{(\alpha)} \widehat{se} \right], \quad (4.1)$$

with $z^{(1-\alpha)}$ and $z^{(\alpha)}$ the relevant normal quantiles. An interval estimate, as given in (4.1), can be more useful than a point estimate $\hat{\theta}$ viewed alone. When the point

estimate and the interval estimate are combined they give an indication of what the “best guess” for θ may be as well as how far that “guess” may be from the actual value of the parameter of interest.

Suppose the data is obtained by random sampling from an unknown distribution,

$$F \rightarrow \mathbf{y} = (y_1, \dots, y_n). \quad (4.2)$$

Let $\hat{\theta} = t(\hat{F})$ be the plug-in estimate of some parameter of interest, $\theta = t(F)$, and let \widehat{se} be an estimate of the standard error of $\hat{\theta}$. For large n , in general we expect the distribution of $\hat{\theta}$ to be approximately normal. Thus, for large n we expect

$$\hat{\theta} \approx N(\theta, \widehat{se}^2), \quad (4.3)$$

or equivalently

$$\frac{\hat{\theta} - \theta}{\widehat{se}} \sim N(0, 1). \quad (4.4)$$

We call the above large-sample, or asymptotic, results and they are often true for general probability models as the size of the sample grows large as well as for statistics other than plug-in estimates [10]. This approximate result gives

$$P \left\{ z^{(\alpha)} \leq \frac{\hat{\theta} - \theta}{\widehat{se}} \leq z^{(1-\alpha)} \right\} \doteq 1 - 2\alpha, \quad (4.5)$$

which can also be written as

$$P \left\{ \hat{\theta} - z^{(1-\alpha)} \widehat{se} \leq \theta \leq \hat{\theta} - z^{(\alpha)} \widehat{se} \right\} \doteq 1 - 2\alpha. \quad (4.6)$$

The approximate $100(1 - 2\alpha)\%$ confidence interval is thus given by

$$\left[\hat{\theta} - z^{(1-\alpha)} \widehat{se}, \hat{\theta} - z^{(\alpha)} \widehat{se} \right]. \quad (4.7)$$

The interpretation of this interval is that on average $100(1 - 2\alpha)\%$ of the time random intervals constructed in this way will contain the true value of θ [10]. It should be remembered that this interval is based on the standard normal approximation and as such gives an approximation to the required coverage probability.

4.2 The Percentile Interval

Let us consider the general situation where we have an observed sample from an unknown probability model, $P \rightarrow \mathbf{y}$, and we generate a bootstrap sample from an estimated probability model, $\hat{P} \rightarrow \mathbf{y}^*$. We calculate the bootstrap replication, $\hat{\theta}^* = \hat{\theta}(\mathbf{y}^*)$ for this bootstrap sample. Recall that $\hat{\theta} \equiv \hat{\theta}(\mathbf{y})$ is the estimator of the parameter of interest, θ . Let \hat{G} be the cumulative distribution function of $\hat{\theta}^*$. The $(1 - 2\alpha)$ percentile interval is then defined by the α th and $(1 - \alpha)$ th percentiles of the cumulative distribution function of the replicates, \hat{G} [10]:

$$[\hat{\theta}_{lo}, \hat{\theta}_{up}] = [\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1 - \alpha)], \quad (4.8)$$

where $\hat{\theta}_{lo}$ represents the lower bound of the interval and $\hat{\theta}_{up}$ represents the upper bound of the interval. The percentile interval can also be written as

$$[\hat{\theta}_{lo}, \hat{\theta}_{up}] = [\hat{\theta}_{\alpha}^*, \hat{\theta}_{1-\alpha}^*], \quad (4.9)$$

where we define $\hat{\theta}_{\alpha}^* \equiv \hat{G}^{-1}(\alpha)$, the $100 \cdot \alpha$ th bootstrap distribution percentile [10]. Equations (4.8) and (4.9) refer to the “ideal bootstrap” where the number of bootstrap replications is infinite, but in practice only a finite number of bootstrap replications are used.

The derivation of the percentile interval through the bootstrap procedure is quite simple. Generate B bootstrap samples from the estimated probability model \hat{P} , $\mathbf{y}_1^*, \dots, \mathbf{y}_B^*$, where B is a large number. For each bootstrap sample a bootstrap replicate, $\hat{\theta}_b^*$, $b = 1, \dots, B$, is calculated. Once the bootstrap replicates, $\{\hat{\theta}_b^*\}$, have been computed for each bootstrap sample, they are sorted in ascending order, $\{\hat{\theta}_{(b)}^*\}$. The α th point of the percentile interval is the $B \cdot \alpha$ th largest value of these sorted replicates. In the same way the $(1 - \alpha)$ th point of the percentile interval is the $B \cdot (1 - \alpha)$ th largest value. In cases where $B \cdot \alpha$ is not an integer let $k = \lfloor (B + 1) \alpha \rfloor$, the largest integer less than or equal to $(B + 1) \alpha$. Then the empirical α and $(1 - \alpha)$ quantiles are the k th and the $(B + 1 - k)$ th largest values of $\{\hat{\theta}_{(b)}^*\}$, respectively. The $(1 - 2\alpha)\%$ bootstrap percentile interval is given by [10]

$$[\hat{\theta}_{lo}, \hat{\theta}_{up}] = [\hat{\theta}_{(\lfloor B\alpha \rfloor)}^*, \hat{\theta}_{(\lfloor B(1-\alpha) \rfloor)}^*]. \quad (4.10)$$

From the central limit theorem it is known that the bootstrap distribution will

be approximately normal as $n \rightarrow \infty$. If the sample size is small, resulting in a very non-normal bootstrap distribution, should the standard (asymptotic) normal interval or the percentile interval be used?

The percentile interval would in general be preferable to the standard interval. The first objection to the use of the standard interval is the normal approximation that underlies it. If n is small this approximation may not be accurate. One way of improving the standard interval is through the use of an appropriate transformation and then mapping the endpoints of the interval back to the original scale. The problem with this approach is that we need to know a different transformation, such as the log-transformation or the exponential-transformation, for each estimator, $\hat{\theta}$, of the parameter of interest, θ . The advantage of the percentile method is that it can be thought of as an algorithm that automatically incorporates these transformations and as a result it extends the effectiveness of the standard interval. In situations where the standard interval would be correct if the appropriate transformation was applied, the percentile method automatically incorporates the transformation and thus we need not know all the appropriate transformations of $\hat{\theta}$; we only need to assume they exist [10]. The percentile interval does not work particularly well in general cases, but in certain cases it is better than the bootstrap-t interval that will be discussed in section 4.3. These include the correlation coefficient. The percentile method also works well for the estimation of quantiles [15].

An advantage of the percentile method should be the improved coverage performance. Although it still tends to under cover, it is more balanced in both sides of the interval than the standard interval. This undercoverage occurs because of the nonparametric inference used. The percentile method has no knowledge of the underlying distribution and uses the empirical distribution instead [10].

A further advantage of this method is that it is transformation respecting. When the interval, obtained after the application of an appropriate transformation on the estimator, $\hat{\theta}$, of the parameter of interest, θ , is mapped back to the original scale, it results in the same interval as before the transformation. This is not the case with the standard interval [10]. The transformation is used to improve the interval and once the endpoints of the interval are transformed back to the original $\hat{\theta}$ scale, it sometimes results in a shorter or longer interval than the interval based on the untransformed estimator. This reflects what is meant by the transformation respecting property. Its result is captured in the percentile interval lemma in [10]:

Suppose the transformation $\hat{\phi} = m(\hat{\theta})$ perfectly normalizes the distribution of $\hat{\theta}$:

$$\hat{\phi} \sim N(\phi, c^2)$$

for some standard deviation c . Then the percentile interval based on $\hat{\theta}$ equals

$$\left[m^{-1}(\hat{\phi} - z^{(1-\alpha)}c), m^{-1}(\hat{\phi} - z^{(\alpha)}c) \right].$$

The percentile interval for any monotone transformation of $\hat{\theta}$ is simply the percentile interval for $\hat{\theta}$ mapped by $m(\hat{\theta})$:

$$[\hat{\phi}_{lo}, \hat{\phi}_{up}] = [m(\hat{\theta}_{lo}), m(\hat{\theta}_{up})]. \quad (4.11)$$

A third advantage of the percentile method is the range-preserving property. Some parameters are defined on a certain range of values, for example the correlation is defined from -1 to 1 . The endpoints of the percentile interval are values of the bootstrap replicates themselves that automatically fall within the allowable range. Confidence procedures that are range-preserving tend to be more accurate and reliable [10].

The percentile interval under complex sampling would be exactly the same as outlined above. The only difference occurs in the calculation of the bootstrap replicates for each sample. Here we make use of the bootstrap weights, w_{hji}^* from (3.30), in the calculation.

4.3 The Bootstrap- t Interval

The bootstrap methodology makes it possible to obtain accurate intervals without making assumptions about approximate normality. The bootstrap- t method considers the “ t -statistic”

$$Z \equiv \frac{\hat{\theta} - \theta}{\widehat{se}}, \quad (4.12)$$

and the approximate confidence interval

$$P\left(\underline{\delta} \leq \frac{\hat{\theta} - \theta}{\widehat{se}} \leq \bar{\delta}\right) = 1 - \alpha,$$

where $\underline{\delta}$ and $\bar{\delta}$ represent, respectively, the lower quantile and upper quantile of the distribution of Z and \widehat{se} is the estimated standard error of $\hat{\theta}$. In the “ideal world” the confidence interval would be

$$P\left(\hat{\theta} - \bar{\delta} \cdot \widehat{se} \leq \theta \leq \hat{\theta} - \underline{\delta} \cdot \widehat{se}\right) = 1 - \alpha,$$

but since we are only considering a sample taken from the population, resampling methods need to be used to estimate the “ideal” situation. The bootstrap-t method uses resampling on the data to generate a bootstrap t-statistic

$$Z^* \equiv \frac{\hat{\theta}^* - \hat{\theta}}{\widehat{se}^*}, \tag{4.13}$$

where $\hat{\theta}^*$ is the statistic calculated on the bootstrap sample, $\hat{\theta}$ is the statistic calculated on the original sample and \widehat{se}^* is the bootstrap standard error. The latter is calculated by resampling from the bootstrap sample, calculating the statistic for each resample and computing the standard error of those bootstrap statistics. Then the bootstrap Z^* value is calculated for each bootstrap sample and ordered, from which the bootstrap quantiles are obtained [10].

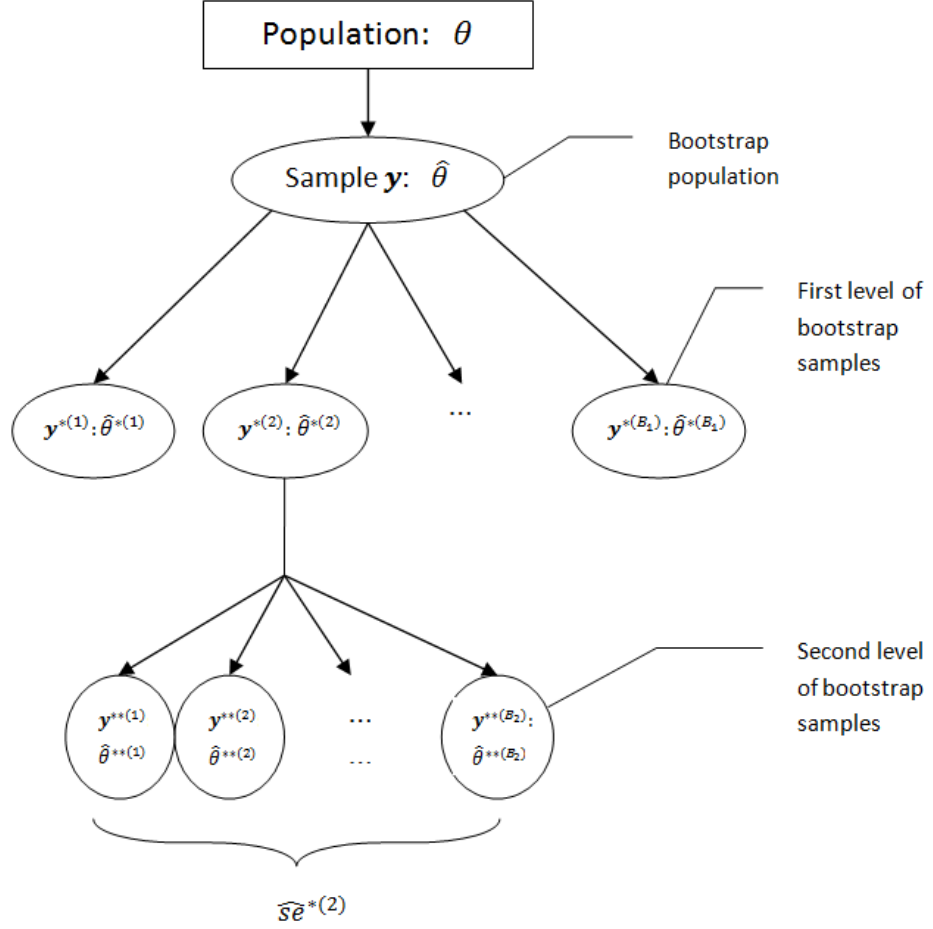


Figure 4.1: Bootstrap- t Confidence Interval

More formally, generate B_1 bootstrap samples $\mathbf{y}_1^*, \dots, \mathbf{y}_{B_1}^*$, with replacement from the original sample and calculate the bootstrap replicate, $\hat{\theta}_b^*$, for each bootstrap sample. B_1 is usually a very large number. For each bootstrap sample we calculate the bootstrap t -statistic

$$t_b^* = \frac{\hat{\theta}_b^* - \hat{\theta}}{\widehat{se}_b^*}, \quad (4.14)$$

where $\hat{\theta}$ is the parameter calculated on the original sample and \widehat{se}_b^* is the estimated standard error of $\hat{\theta}_b^*$ for the bootstrap sample \mathbf{y}_b^* [10]. This estimated standard error is obtained by taking B_2 bootstrap samples from the current bootstrap sample, \mathbf{y}_b^* , calculating the replicates for each resample, $\{\hat{\theta}_b^{**}, b = 1, \dots, B_2\}$, and then obtaining the standard error of those replicates. We distinguish here between firstly

taking B_1 bootstrap samples from the original sample and then taking B_2 resamples from each of the B_1 bootstrap samples to emphasize the use of a nested bootstrap. Both B_1 and B_2 are typically large, but need not be the same size. The α th percentile of t_b^* , $b = 1, \dots, B_1$, is estimated by the value $\hat{t}^{(\alpha)}$ such that

$$\frac{\#\{t_b^* \leq \hat{t}^{(\alpha)}\}}{B_1} = \alpha. \quad (4.15)$$

The $\{t_b^*\}$ values are sorted in ascending order and the α th point is the $B_1 \cdot \alpha$ th largest value of these $\{t_{(b)}^*\}$ values. In the same way the $(1 - \alpha)$ th point is the $B_1 \cdot (1 - \alpha)$ th largest value. In cases where $B_1 \cdot \alpha$ is not an integer let $k = [(B_1 + 1)\alpha]$, the largest integer less than or equal to $(B_1 + 1)\alpha$. Then the empirical α and $(1 - \alpha)$ quantiles are the k th largest and the $(B_1 + 1 - k)$ th largest values of $\{t_{(b)}^*\}$, respectively [10]. The $(1 - 2\alpha)\%$ bootstrap-t interval for θ is then given by

$$\left(\hat{\theta} - \hat{t}^{(1-\alpha)} \cdot \widehat{se}_{B_1}, \hat{\theta} - \hat{t}^{(\alpha)} \cdot \widehat{se}_{B_1} \right), \quad (4.16)$$

where \widehat{se}_{B_1} is the estimated standard error of $\hat{\theta}$ calculated as the standard error of $\{\hat{\theta}_b^*, b = 1, \dots, B_1\}$. It should be noted that $B_1 = 100$ or 200 is not adequate for the construction of confidence intervals. Many more bootstrap samples are required to accurately estimate the parameter of interest, θ , according to the argument of Booth and Sarkar [3] briefly referred to in section 3.2.3, and then there is a second level of bootstrapping needed to estimate the standard error of each bootstrap replicate, $\hat{\theta}_b^*$.

It has been shown that the coverage of the bootstrap-t interval tends to be closer to the desired level than that of the standard interval. Unfortunately the gain in accuracy goes hand in hand with a loss in generality, since the bootstrap-t interval applies only to the given sample [10].

The interval generated by the bootstrap-t method is not symmetric about zero. It is this asymmetry that plays an important part in the coverage improvement that is enjoyed by the bootstrap-t. This method is particularly applicable to location statistics such as the mean, median, trimmed mean or sample percentile. When it comes to more general problems such as generating a confidence interval for the correlation coefficient, the bootstrap-t method cannot be trusted [10].

There is a major computational difficulty with the use of the bootstrap-t interval. It requires the computation of the standard error of each $\hat{\theta}_b^*$ where $b = 1, \dots, B_1$. The standard error, \widehat{se}_b^* , has to be estimated for each bootstrap sample which is not a

problem when the parameter of interest is the sample mean, because there exists a formula for its standard error. Unfortunately there exists very few standard error formulas which means that the standard error for other statistics will have to be estimated and this leads to a nested bootstrap. Thus, in a nested bootstrap where B_1 bootstrap samples are taken from the original sample and B_2 resamples are taken from each of the B_1 bootstrap samples to estimate the standard error, lead to $B_1 \cdot B_2$ bootstrap samples. This is a large number and hence computationally intensive. Given this difficulty with the computational demand of the bootstrap-t method, the $\{\widehat{se}_b^*\}$ will rather be estimated using the jackknife method.

Now consider the application of the bootstrap-t interval in complex sampling. In [25], Rao and Wu obtained these intervals for smooth functions, $\hat{\theta} = \hat{\theta}(\mathbf{y})$, by approximating the distribution of

$$Z = \frac{(\hat{\theta} - \theta)}{\sqrt{\hat{V}_{JK}(\hat{\theta})}}, \quad (4.17)$$

through the use of the bootstrap method. Recall that $\hat{\theta}$ is the estimator of the population parameter θ and $\hat{V}_{JK}(\hat{\theta})$ is the jackknife estimate of the variance of $\hat{\theta}$ as given in (3.11). The bootstrap counterpart is given by

$$Z^* = \frac{(\hat{\theta}^* - \hat{\theta})}{\sqrt{\hat{V}_{JK}(\hat{\theta}^*)}}, \quad (4.18)$$

where $\hat{V}_{JK}(\hat{\theta}^*)$ is similar to the jackknife estimate of the variance of $\hat{\theta}$ under complex sampling in (3.16), but is the estimated variance of the jackknife replicates calculated from the second level of B_2 bootstrap samples. Some literature recommend that the weights (3.13) in the jackknife formulas given in section 3.1.2,

$$w_{i(hj)} = \begin{cases} w_i, & i \notin h \\ w_i \cdot \frac{n_h}{(n_h-1)}, & i \in h, i \notin j \\ 0, & i \in (h, j) \end{cases},$$

are simply replaced by the bootstrap weights w_{hji}^* [26] given by

$$w_{hji}^* = w_{hji} \left[\left\{ 1 - \sqrt{\left(\frac{m_h}{n_h - 1} \right)} \right\} + \left(\sqrt{\frac{m_h}{n_h - 1}} \right) \left(\frac{n_h}{m_h} \right) m_{hj}^* \right], \quad (4.19)$$

where m_{hj}^* is the number of times that the (hj) -th sample psu is selected for the specific bootstrap sample. These are the bootstrap weights calculated on the first level. Since this recommendation raises some doubt on our part, the jackknife weights will be recalculated on the second level and used in the calculation of the jackknife replicates. For every bootstrap sample, calculate $t_b^*, b = 1, \dots, B$ and sort the values in ascending order. The two-sided $(1 - 2\alpha)\%$ bootstrap-t confidence interval under complex sampling is then given by

$$\left[\hat{\theta} - t_U^* \cdot \hat{S}_B(\hat{\theta}), \hat{\theta} - t_L^* \cdot \hat{S}_B(\hat{\theta}) \right], \quad (4.20)$$

where t_L^* and t_U^* are the lower and upper α -points obtained from $t_{(1)}^*, \dots, t_{(B_1)}^*$

$$t_b^* = \frac{(\hat{\theta}_b^* - \hat{\theta})}{\sqrt{V_{JK}(\hat{\theta}_b^*)}}, \quad b = 1, \dots, B_1.$$

$\hat{S}_B(\hat{\theta})$ is the bootstrap estimated standard error of $\hat{\theta}$ and $\hat{V}_{JK}(\hat{\theta}_b^*)$ is the jackknife estimate of variance calculated from the second level of bootstrapping [26] as previously described for a SRS. They, namely t_L^* and t_U^* , correspond to the $B_1 \cdot \alpha$ th and the $B_1 \cdot (1 - \alpha)$ th largest values of the sorted $\{t_b^*\}$ values. If $B_1 \cdot \alpha$ is not an integer, the same argument can be followed as given above. Also in complex sampling, a variance stabilizing transformation can be used to correct uneven error rates, but the bootstrap provides an alternative when such transformations do not exist or are unknown [26].

The bootstrap methodology provides a good measure for both smooth and non-smooth functions. It is the preferred method for one-sided intervals, but if suitable variance-stabilizing transformations can be found then other methods, such as the normal-theory one-sided interval, may be used and may perform better. As it is generally difficult to find these transformations the bootstrap intervals will be used [15].

4.4 Conclusion

In this chapter confidence intervals were discussed. A brief overview of the standard (asymptotic) interval was given followed by discussions on the bootstrap-t interval and the percentile interval. Their performance was compared to the standard interval and advantages and disadvantages were discussed for each resampling confidence interval method relative to the standard interval .

The percentile method is generally a preferred method to the standard interval. Its virtues include better coverage performance, transformation respecting as well as range-preserving. Methods that have these properties tend to be more accurate and reliable than other methods. It is also a very simple method to implement. Its disadvantage is that it does not work well in general.

The bootstrap methodology makes it possible to obtain accurate intervals without making assumptions about approximate normality. The bootstrap-t interval's coverage is closer to the desired level, 2α , and it provides good measures for both smooth and non-smooth functions. A disadvantage that was mentioned is its erratic performance in small-sample, nonparametric settings. The data that will be used in the simulation study is very large which should invalidate this problem. Another disadvantage that most certainly will still be problematic is the nested bootstrap that forms part of the computation of the bootstrap-t interval. Recall that it requires $B_1 \cdot B_2$ resamples where B_1 and B_2 are both large. It is thus computationally intensive, especially for large samples.

Although the percentile method certainly has a number of advantages as well as properties that the bootstrap-t interval does not have, it is perhaps too simple. Nevertheless, both the percentile method and the bootstrap-t method will be used in the construction of confidence intervals and the results will then be used to compare these methods.

Chapter 5

Indicators of Poverty and Inequality

5.1 Measures of poverty

According to the World Bank (2000), “poverty is pronounced deprivation in well-being.” [11]

Well-being can be seen as the command over commodities in general [11]. People are better off if they have a greater command over resources. The question is whether individuals or households have enough resources to meet their needs. Poverty is measured as a comparison of an individual’s income or expenditure, where some threshold is defined, below which individuals are considered to be poor or to have pronounced deprivation is well-being. The poverty line can be seen as this threshold and is used as benchmark to determine whether a household is “in poverty” [11]. There can be distinguished between

- absolute poverty, and
- relative poverty.

Relative poverty refers to the poorest segment of the population whereas an absolute poverty line is set to represent the same purchasing power every year. Relative poverty is useful for groups or organisations dedicated to helping the poor. Absolute poverty lines are necessary when the intent is to determine the effectiveness of anti-poverty policies or when poverty rates between different countries need to be compared [11].

The common poverty lines used by the World Bank are the one dollar per day and two dollars per day poverty lines.

Poverty can also be seen as an individual's lack of capability of obtaining a specific level of consumption of goods such as food, shelter, education, etc. [11]. By this view there has to be looked beyond the traditional monetary measures of poverty.

Lastly, the broadest approach to well-being is the argument that it comes from a "capability" to function in society [11]. This implies that poverty occurs when people lack capabilities which lead to inadequate income, education, poor health, etc. In this view poverty is a multi-dimensional phenomenon that is less agreeable to easy solutions.

Why should we measure poverty [11]?

1. To keep the poor on the agenda:

The poor can easily be ignored which makes the measurement necessary if it is to appear on the political and economic agenda.

2. To target interventions, domestically and worldwide:

Interventions are targeted because the poor cannot be helped if we do not at least know who they are.

3. To monitor and evaluate projects and policy interventions geared towards the poor:

This is done to be able to predict the effects of, and evaluate, policies and programs designed to help the poor. Policies sometimes look good on paper but do not work as well in practice.

4. To evaluate the effectiveness of institutions whose goal it is to help the poor:

Accurate information on poverty is essential for knowing if a government is doing an acceptable job of combating poverty.

The following poverty indices will be considered in this thesis:

1. Headcount Index.
2. Poverty Gap Index.
3. Squared Poverty Gap Index.

5.1.1 Headcount Index

This is undoubtedly the most widely-used measure of poverty [11]. It measures the proportion of the population that is counted as being poor:

$$P_0 = \frac{N_p}{N}, \quad (5.1)$$

where N_p is the number of people measured as poor and N is the total population size. Another, sometimes more useful, form of the headcount index is

$$P_0 = \frac{1}{N} \sum_{i=1}^N I(y_i < z). \quad (5.2)$$

$I(\cdot)$ represents the indicator function, y_i is the per capita income of person i and z is the poverty line, i.e.

$$I(y < z) = \begin{cases} 1, & y < z \\ 0, & \text{otherwise} \end{cases}. \quad (5.3)$$

In a complex sample, as is usually the case in practice, the formula would incorporate weights. Thus, the headcount index expressed for a complex sample is given by

$$P_H = \frac{1}{\sum_{i \in S} w_i} \sum_{i \in S} w_i I(y_i < z), \quad (5.4)$$

where w_i is the weight associated with person i and $i \in S$ refers to all observations in the sample [5]. Recall that the weight associated with a person is the number of population elements represented by that person. Thus, $\sum_{i \in S} w_i$ represents the population size, N , and the weight multiplied with the indicator function equals the number of persons in the population that are poor. This again gives the same formula as for a simple random sample.

One advantage of this measure is that it is simple to construct and easy to understand. However, it has a few shortcomings. Firstly, it does not take the intensity of the poverty into account. The headcount index as a welfare function violates the transfer principle. The transfer principle refers to the transfer of riches from a richer person to a poorer person that should improve the measure of welfare of the latter. Secondly, it does not indicate the level of poverty of the poor. This

implies that the measure doesn't change when the people below the poverty line become more poor. Lastly, poverty estimates should be calculated for individuals and not households. Most survey data are related to households, with the result that for a measure of poverty at individual level, we have to make the assumption that all members of a given household enjoy the same level of well-being. This may not always be the case.

5.1.2 Poverty Gap Index

This measure adds up the extent to which individuals on average fall below the poverty line. It expresses this extent as a percentage of the poverty line [11]. The poverty gap, G_i , is defined as the poverty line less the actual income of the i th poor individual. The gap is considered to be zero for all other individuals if the difference is negative.

$$G_i = (z - y_i)I(y_i < z). \quad (5.5)$$

This gives the poverty gap index as

$$P_1 = \frac{1}{N} \sum_{i=1}^N \frac{G_i}{z}, \quad (5.6)$$

which could be seen as the “gap” between the poverty line z and the per capita income of person i as defined in (5.5). The poverty gap index once again makes use of weights in a complex sample:

$$P_1 = \frac{1}{\sum_{i \in S} w_i} \sum_{i \in S} w_i \frac{G_i}{z}, \quad (5.7)$$

where w_i is the weight associated with person i [5].

Sometimes it helps to think of this measure as the cost of eliminating poverty because it shows how much must be transferred to the poor in an attempt to bring their income up to the poverty line.

5.1.3 Squared Poverty Gap Index

This is also referred to as the poverty severity index. The squared poverty gap index is a measure that takes into account inequality among the poor. It follows

that by squaring the poverty gap index the measure implicitly puts more weight on observations that fall well below the poverty line [11]. Define

$$P_2 = \frac{1}{N} \sum_{i=1}^N \left(\frac{G_i}{z} \right)^2, \quad (5.8)$$

with G_i defined as in (5.5).

Squared poverty gap index in complex sampling:

$$P_2 = \frac{1}{\sum_{i \in S} w_i} \sum_{i \in S} w_i \left(\frac{G_i}{z} \right)^2, \quad (5.9)$$

where w_i is the weight associated with person i [5].

Unfortunately this measure lacks intuitive appeal. It is not easy to interpret and not widely used. One could consider a more general form of the measure

$$P_\alpha = \frac{1}{N} \sum_{i=1}^N \left(\frac{G_i}{z} \right)^\alpha, \quad (5.10)$$

where a measure of the sensitivity, α , of the index to poverty and the poverty line is introduced [11]. If $\alpha = 1$, then we have the poverty gap index and for $\alpha = 2$, we have the squared poverty gap index. If $0 < \alpha < 1$, the measure decreases in the living standard of the poor and for $\alpha > 1$ the increase in measured poverty due to the fall in an individual's living standard will be greater the poorer the individual is [11]. The extension of this measure to complex sampling would incorporate the weights as in the squared poverty gap index. A major problem is that the choice of the best value for α remains unsolved.

5.2 Measures of Inequality

Inequality is a much broader concept than poverty because it is defined over the entire population, not only for the population below a certain poverty line [11]. One of the desirable properties of inequality measures is their independence from the mean of a distribution. The easiest way to measure inequality is by dividing the population into five levels ranging from poorest to richest. At each level the proportion of income accumulated is reported [11].

The following are criteria that make a good measure of income inequality [11]:

1. Mean independence:

This implies that if all incomes were doubled \therefore mean shift a lot, the inequality measure remains the same.

2. Population size independence:

If the population size were to change, the measure of inequality should not change, *ceteris paribus*.

3. Symmetry:

If two people swop income then there should be no change in the measure of inequality.

4. Pigou-Dalton transfer sensitivity:

The transfer of income from rich to poor reduces the measured inequality.

5. Decomposability:

Inequality can be broken down by population groups.

6. Statistical testability:

It allows the test of significance of changes in the index over time.

5.2.1 Gini Coefficient of Inequality

This is the most widely used measure of inequality. It is based on the Lorenz curve which is a cumulative frequency curve. The curve compares the distribution of a specific variable to the uniform distribution that represents equality [11]. The coefficient is constructed by graphing the cumulative percentage of households, from poor to rich, on the horizontal axis and the cumulative percentage of income or expenditure on the vertical axis.

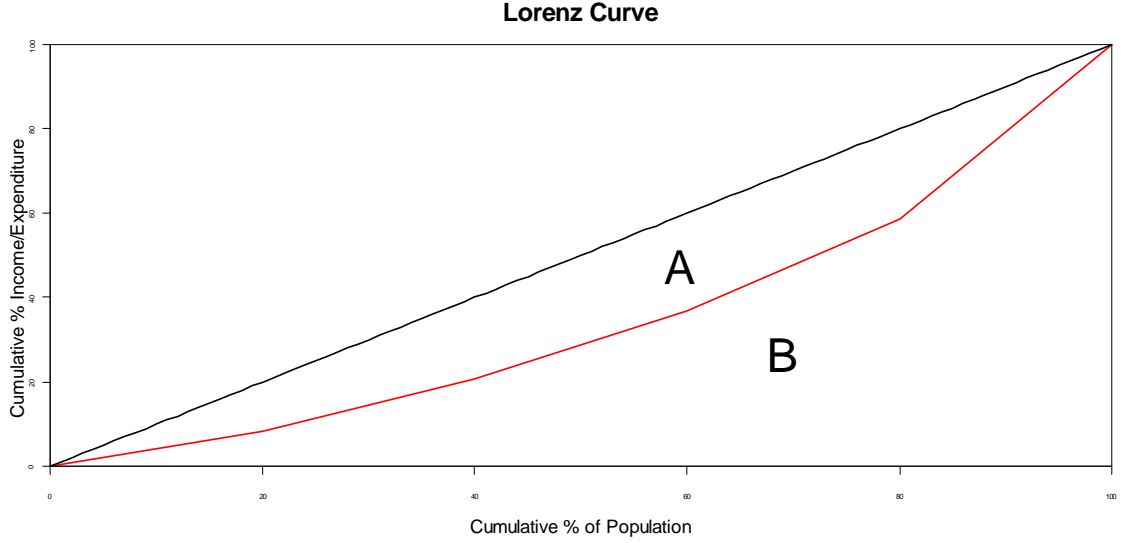


Figure 5.1: Lorenz Curve

Figure 5.1 was constructed from the Vietnamese data in the World Bank Institute August 2005 report on Poverty Analysis. The diagonal line represents perfect equality. The Gini coefficient is then calculated as

$$Gini = \frac{A}{(A + B)} = 2A, \quad (5.11)$$

where A and B are the indicated areas in fig. 1.5. When $A = 0$, then Gini becomes zero and we have perfect equality. If $B = 0$, then Gini becomes one and we have perfect inequality. If we let x_i be a point on the x -axis and y_i be a point on the y -axis, then we get a formal definition of the Gini coefficient [11]:

$$Gini = 1 - \sum_{i=1}^N (x_i - x_{i-1}) (y_i + y_{i-1}), \quad (5.12)$$

and if there are N equal intervals on the x -axis, then the formula simplifies to

$$Gini = 1 - \frac{1}{N} \sum_{i=1}^N (y_i + y_{i-1}). \quad (5.13)$$

In a complex sample the formula becomes [5]

$$Gini = \frac{2}{\sum_{i \in S} w_i y_i} \sum_{i \in S} w_i \hat{F}(y_i) y_i - 1, \quad (5.14)$$

where y_i is the per capita income of person i and

$$\hat{F}(y) = \frac{1}{\sum_{i \in S} w_i} \sum_{i \in S} w_i I(y_i \leq y), \quad (5.15)$$

is an estimate of the income distribution function [11].

Although it is the most widely used measure it is not entirely satisfactory. It is not easily decomposable so that the total Gini of society is equal to the sum of the Gini's of its subgroups.

5.2.2 Generalized Entropy Measures

This set of measures satisfies all six of the above criteria for a good measure of inequality. The most widely used measures among these are the Theil indices [11]. The general formula for these generalized entropy measures is

$$GE(\alpha) = \frac{1}{\alpha(\alpha - 1)} \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{y_i}{\bar{y}} \right)^\alpha - 1 \right], \quad (5.16)$$

where \bar{y} is the mean income. These measures vary in size ranging from zero to ∞ . The closer to zero, the closer to an equal distribution of income. The closer the value to ∞ , the higher the level of inequality. Note that the parameter α in this case does not refer to the measure of sensitivity in the squared poverty gap index, but rather represents the weight associated with the distance between incomes at different levels of the income distribution [11]. The lower the value of α , the more sensitive the measure is to changes in the lower tail of the income distribution and the greater the value of α , the more sensitive the value is to changes in the upper tail. The values used most often are $\alpha = 0, 1, 2$. For $\alpha = 0$:

$$GE(0) = \frac{1}{N} \sum_{i=1}^N \ln \left(\frac{\bar{y}}{y_i} \right), \quad (5.17)$$

known as Theil's L index. It is sometimes referred to as the mean log deviation (MLD) measure [11]. For $\alpha = 1$:

$$GE(1) = \frac{1}{N} \sum_{i=1}^N \frac{y_i}{\bar{y}} \ln \left(\frac{\bar{y}}{y_i} \right), \quad (5.18)$$

and is called Theil's T index. Theil's T index is more sensitive to change in the

higher income groups than the MLD. Its application in complex sampling gives the following formulas:

$$GE_w(0) = \frac{1}{\sum_{i \in S} w_i} \sum_{i \in S} w_i \ln \left(\frac{\bar{y}_w}{y_i} \right) \quad (5.19)$$

and

$$GE_w(1) = \frac{1}{\sum_{i \in S} w_i} \sum_{i \in S} w_i \frac{y_i}{\bar{y}_w} \ln \left(\frac{y_i}{\bar{y}_w} \right), \quad (5.20)$$

where \bar{y}_w is the weighted mean per capita income [5].

5.2.3 Atkinson's Inequality Measures

Another class of inequality measures that is becoming increasingly popular to use, is Atkinson's inequality measures [11]. It measures the cumulative deviation from the “equally distributed equivalent income”, which is the “level of income per head which if equally distributed would give the same level of social welfare as the present distribution” [28]. This class makes use of a parameter, ε , which reflects society's aversion to inequality [2] and the Atkinson measures have similar theoretical properties to those of the Gini index. This parameter can take on any values from zero to infinity and can be interpreted as follows [2] [28]:

- For $\varepsilon = 0$ the income transfers have weights equal to zero and the distributions are ranked simply in terms of total level of income;
- For $\varepsilon > 0$, there exists an aversion to inequality;
- As ε grows large, the more weight society attaches to income transfers at the lower end of the distribution;
- When ε is close to infinity, the inequality is only sensitive to income transfers at the lowest levels of the distribution.

Typical values for ε are 0.5 and 2 [2]. In general the measures are defined as

$$A_\varepsilon = \begin{cases} 1 - \left[\frac{1}{N} \sum_{i=1}^N \left(\frac{y_i}{\bar{y}} \right)^{1-\varepsilon} \right]^{1/(1-\varepsilon)}, & \varepsilon \neq 1 \\ 1 - \frac{\prod_{i=1}^N \left(y_i^{(1/N)} \right)}{\bar{y}}, & \varepsilon = 1 \end{cases},$$

where y_i is the income (or alternatively, the equivalent income) as defined in (5.22) of person i and \bar{y} is the mean income. In a complex sample the resulting formula would be

$$A_\varepsilon = \begin{cases} 1 - \left[\frac{1}{\sum_{i \in S} w_i} \sum_{i \in S} w_i \left(\frac{y_i}{\bar{y}} \right)^{1-\varepsilon} \right]^{1/(1-\varepsilon)}, & \varepsilon \neq 1 \\ 1 - \frac{\prod_{i \in S} w_i \left(y_i^{1/(\sum_{i \in S} w_i)} \right)}{\sum_{i \in S} w_i y_i}, & \varepsilon = 1 \end{cases}, \quad (5.21)$$

where w_i is the weight associated with person i .

Remark: Although the Atkinson inequality measures are discussed here, they do not form part of the simulation study. They are only discussed for completeness.

5.3 Laeken Indicators

In December 2000, at the Nice European Council, Heads of State and Government of the European Union reconfirmed and implemented their decision made in Lisbon, March 2000, that the fight against poverty and social exclusion would be best achieved through an open method of co-ordination [12]. This method of co-ordination has three key elements:

1. The definition of commonly-agreed objectives for the European Union (EU) as a whole.
2. The development of national action plans to meet these commonly-agreed objectives.
3. The periodic reporting and monitoring of the progress made with these action plans.

In December 2001, the Laeken European Council authorized a first set of eighteen indicators for social inclusion [16]. They include the Gini coefficient and the at-risk-of-poverty rate measures. These indicators will allow monitoring of Member States' progress in achieving the EU objectives in a comparable manner. The Laeken indicators cover four important aspects of social inclusion:

1. Financial poverty.
2. Employment.

3. Health.

4. Education.

These dimensions for indicators give a good representation of the social concerns of the EU. In order to study these social concerns effectively, the Laeken indicators as a group need to be considered. However, due to time constraints, we shall only consider the following indicators with regard to the financial dimension:

- at-risk-of poverty rate, I_1 ;
- indicators based on quantiles including the median, I_{1e} , and the quintile share ratio, I_2 ;
- the Gini coefficient, I_{14} .

A key feature in the definition of Laeken indicators, is equivalized disposable income [16]. This is defined as a new variable called *eqinc* that is associated with an individual. Let person i be in household k . Then the equivalized income for that individual is calculated as

$$eqinc_i = \frac{totinc_k}{eqsize_k}. \quad (5.22)$$

It states that the equivalized income is obtained by dividing the total income of the household that person belongs to by the equivalized size of that household. The equivalized size is obtained from the sum of personal OECD (Organisation for Economic Co-operation and Development) weights [16]. These weights for household k are defined as $[1, 0.5, 0.3]$ where 1 is assigned to the first adult, 0.5 to all other adults aged 14 and older and 0.3 to all children younger than 14 [16].

5.3.1 Definitions of Indicators

At-risk-of-poverty rate

This is the share of people that have an equivalized total net income below 60% of the national median income [12]. After the equivalized income has been calculated for each person they are sorted in ascending order according to their equivalized income. The median is then the equivalized income of the person for whom the cumulative sum of person weights is less than or equal to 50% of the total sum of

weights [12]. Due to this view, persons in the same household are placed together on the same side of the median. Once the median income has been established, the at-risk-of-poverty threshold is calculated as 60% of this median.

Inequality of income distribution S80/S20 quintile share ratio

It is defined as the ratio of equivalized total net income received by 20% of the population with the highest income (top quintile) to the equivalized total net income received by 20% of the population with the lowest income (lowest quintile) [12]. To be able to calculate this ratio the population has to be grouped into quintiles. Firstly the individuals are sorted in ascending order according to the equivalized income. The 20% of individuals at the lower end of the distribution represent the 20% of individuals defined as poorest (first quintile). The 20% at the upper end of the distribution represent the 20% defined as the 20% richest individuals (fifth quintile). The cut-off point for individuals to be part of a specific quintile is set in such a way that the cumulated sum of weights is less than or equal to $x \cdot 20\%$ of the total sum of weights where $x = 1, \dots, 5$. [12]

The equivalized total net income of a specific quintile is equal to the sum of the equivalized income of individuals belonging to that quintile. The mean of the equivalized income of the quintile is, in practice, preferred to the equivalized total net income for use in the calculation of the quintile share ratio. The mean is used to minimise the impact from the fact that the number of persons in a quintile can vary from the expected 20% of the total population during the quintile distribution process [12].

Inequality of income distribution: Gini coefficient

This indicator is defined as the relationship of cumulative shares of the population, arranged according to the level of income to the cumulative share of the equivalized total net income received by them [12]. It is used to measure the level of inequality in an income distribution [16]. Once again the persons are sorted according to their equivalized total net income. All persons with unknown equivalized income are excluded. The formula for calculating the Gini coefficient as The Eurostat document [12] writes,

$$GINI = 100 \times \left[\frac{2 \sum_{i=1}^{last\ person} \left(w_i eqinc_i \sum_{j=1}^{person\ i} w_j \right) - \sum_{i=1}^{last\ person} (w_i)^2 eqinc_i}{\left(\sum_{i=1}^{last\ person} w_i \right) \sum_{i=1}^{last\ person} (w_i eqinc_i)} - 1 \right]. \quad (5.23)$$

Equation (5.14) can be seen as an approximation to (5.23), with y_i replaced by the new variable, $eqinc_i$.

5.3.2 Estimation of Indicators

In order to define the estimators for specific indicators, we first give a general definition of sample quantiles. For a sequence $\{x_i\}$, let $\{x_{(i)}\}$ denote the corresponding ordered sequence, and w_i a weight corresponding to $x_{(i)}$. For $0 < \alpha < 1$, the α -th (sample) quantile is defined as

$$q_\alpha = \begin{cases} \frac{(x_{(j)} + x_{(j+1)})}{2}, & \text{if } \sum_{i=1}^j w_i = \alpha \hat{N} \\ x_{(j+1)}, & \text{if } \sum_{i=1}^j w_i < \alpha \hat{N} < \sum_{i=1}^{j+1} w_i \end{cases}, \quad (5.24)$$

where $\hat{N} = \sum_{i \in S} w_i$.

For example, $\alpha = 0.5$ gives the median and $\alpha = 0.2, 0.4, 0.6, 0.8$ give the quintiles [16].

Remark:

In our applications, $\{w_i\}$ refers to the number of people in the population represented by the i -th individual and thus \hat{N} is an estimator of the population size.

At-risk-of-poverty rate (I_1)

From the definition given above, the at-risk-of-poverty rate can be formulated as

$$I_1 = \frac{\sum_i w_i I(x_{(i)} \leq threshold)}{N},$$

where the threshold is defined as 60% of the median income of the population and N is the size of the population.

The estimation of I_1 makes use of the quantile estimators as defined above. The at-risk-of-poverty threshold is estimated by

$$\hat{I}_{1e} = 0.6q_{0.5}. \quad (5.25)$$

where $q_{0.5}$ is the estimated median income obtained from the above quantile estimators for α equal to 0.5 . Once the threshold has been estimated the at-risk-of-poverty indicator becomes [12]

$$\hat{I}_1 = \frac{\sum_{i: eqinc < threshold} w_i}{\sum_{i \in S} w_i}. \quad (5.26)$$

Income Quintile Share Ratio (I_2)

The quintile share ratio is defined as

$$I_2 = \frac{S_{80}}{S_{20}} = \frac{\sum_{i \in 5th \text{ quintile}} w_i * eqinc_i}{\sum_{i \in 1st \text{ quintile}} w_i * eqinc_i}. \quad (5.27)$$

This indicator is estimated as

$$\hat{I}_2 = \frac{\sum_{i \in S: eqinc > q_{0.8}} w_i * eqinc_i}{\sum_{i \in S: eqinc < q_{0.2}} w_i * eqinc_i}, \quad (5.28)$$

where $q_{0.2}$ and $q_{0.8}$ are the estimated first and fifth quintiles of the weighted equivalized income of individuals in the sample, as defined in (5.22) [16].

Gini Coefficient

A great deal of the understanding of the Gini coefficient lies in understanding the meaning of the Lorenz curve. The connection between the Gini coefficient and the Lorenz curve has been illustrated above. In the case of the Laeken indicators, sort the equivalized income in ascending order. Let N be the size of the population and $Y_i = \sum_{j=1}^i y_{(j)}$, where Y_i represents the total income earned by the first i members of the income-sorted population [16]. For convenience we denote $Y_0 = 0$. From this view the Lorenz curve is simply a graph of Y_i against i , where $i = 0, \dots, N$. The Gini coefficient, as one of the Laeken indicators, has the same definition with regard to the Lorenz curve as previously described. Two populations can then be compared with regard to the values of their respective Gini coefficients. The population with the higher Gini coefficient will have greater inequality [16]. Another interpretation of the Gini coefficient is that it can be seen as the expected income gap between

two randomly selected individuals from the population [16]. The formal definition for the estimator of the Gini coefficient is given by

$$\hat{G} = \frac{2 \sum_{i \in S} w_i y_{(i)} \hat{N}_i - \sum_{i \in S} w_i^2 y_{(i)}}{\hat{N} \sum_{i \in S} w_i y_{(i)}} - 1 \quad (5.29)$$

where $\hat{N} = \sum_{i \in S} w_i$ is the estimated population size and $\hat{N}_i = \sum_{j=1}^i w_j$ is the estimated population size with equivalized income less than, or equal to, that of person i . Note that w_i is the weight that corresponds to the sorted equivalized income of person i , $y_{(i)}$.

5.4 Conclusion

In this chapter measures of poverty and inequality were discussed. The two poverty lines to be used in this study for the estimation of the poverty measures were also introduced, namely an absolute poverty line and a relative poverty line. Two income variables that will be used for the estimation of the inequality measures, are *percy*, the ratio of total household income to household size, and equivalized income (*eqinc*), the ratio of total household income to equivalized household size. All the measures described in this chapter will be included in the simulation study, except the Atkinson measures. They were included simply for completeness.

Chapter 6

Description of the Data

The previous chapters gave an overview of complex sampling, variance and confidence interval estimation through the use of resampling techniques as well as measures of income and inequality. The present chapter marks the beginning of the practical implementation of the theory and techniques discussed thus far. It will contain a description of the dataset, the Income and Expenditure Survey 2005, as well as how Statistics South Africa conducted the survey. Aspects of the survey that will receive attention include the design and the weighting used.

6.1 Income and Expenditure Survey 2005/2006

The dataset that will be used in the analysis and that will act as surrogate population, is the Income and Expenditure survey conducted over the period September 2005 until August 2006, hereafter referred to as IES 2005/2006. The intention of this survey is to examine income and expenditure as well as poverty and inequality in South Africa. Households that were sampled took part in the survey for one month after which new subsamples of households started taking part in the survey at the beginning of each month [1].

For this most recent IES, Statistics SA changed the methodology used in previous surveys of this kind. Previously the recall method was used, but now a combination of the recall method and the diary method was used. In a nutshell, a main questionnaire consisting of five interview modules is administered by a fieldworker to a selected household. Each interview was conducted on five different visits. The main questionnaire required households to account for their acquisitions of the following

goods and services [1]:

- Durable

Items or services that last a long time. For example cars, furniture, etc.

- Semi-durable

Items that require replacement more often than durable items. For example clothing, shoes, etc.

This information, as well as income acquired by different members of a household, was collected over the eleven months prior to the survey.

The new part of the survey methodology required households to keep a diary of their daily acquisitions over the four weeks of the survey. These diaries were collected on a weekly basis and the purpose was to ensure that the information collected was as close as possible to the period of transaction. Information collection was based on acquisition that takes into account the total value of all goods and services acquired during a given period [1].

6.1.1 Data Collection Methods

Three methods were used to collect the survey information [1]:

1. Main Questionnaire

It consisted of a booklet of questions administered to respondents during the course of the survey month. As mentioned before, the main questionnaire consisted of five parts. The first part covered household characteristics, the next three parts covered different categories of consumption expenditure and the final part covered household income.

2. Weekly Diary

Each household had to write down their daily acquisitions according to specific “categories” namely the nature, type, source and purpose of the item acquired.

3. Summary Questionnaire

The fieldworker had to “summarize” the total value of each item acquired during the week and then had to transfer it to the appropriate section of the

questionnaire. This assisted the fieldworker in summarising the consumption expenditure of each household during the survey month.

6.1.2 Response and Imputation of non-response

As discussed in chapter two, there are two types of non-response, namely unit non-response and item non-response. Unit non-response is taken care of during weighting while item non-response requires imputation at different levels. Here, imputation was done for missing diaries as well as item non-response [1].

An imputation method called cell mean imputation is used by Statistics SA. This method divides the data into groups according to variables with no missing values. The mean value is then imputed into the missing values [18]. For the missing diaries, households were divided into groups according to the number of diaries completed within the four weeks of the survey. Those households with less than two diaries or a diary but no main questionnaire were considered nonrespondent. The mean expenditure of respondent households, those with two or more completed diaries, were imputed [1].

For the item non-response, respondent households with similar characteristics to the nonrespondents were grouped together and the average value for these households were imputed [1].

6.2 Survey Design

The sampling frame used for IES 2005/2006 was a newly designed master sample based on the enumeration areas of the 2001 population census [1]. The selection of psu's require the availability of a frame or list of all psu's. When such a frame is used for multiple surveys or multiple rounds of the same survey, it is known as a master sample frame. A master sample is a sample from which subsamples can be selected to serve the needs of more than one survey or survey round [24]. An enumeration area (EA) is the smallest geographical unit (piece of land) into which the country is divided for survey purposes and EA's were used as psu's. The master sample consists of all households in South Africa while the target population consists of all eligible persons and households in the country. The master sample considered all households living in private dwelling units as well as workers living in workers'

quarters. A dwelling unit (DU) is defined as any structure or part of a structure or group of structures occupied or meant to be occupied by one or more than one household [1].

The 3000 EA's in the master sample were stratified into four groups of 750 EA's each. The EA's were used as psu's. A random sample of 250 psu's were selected each month. From each selected psu, a systematic sample of 8 DU's was chosen. Thus, a stratified two-stage cluster sample was used with the four groups as explicit stratification variable, enumeration areas as psu's and dwelling units as ssu's. So, 24000 DU's were interviewed over the twelve month period. This design ensured that the sample was evenly spread over the twelve months while being nationally representative in each of the four groups [1].

6.3 Weighting

Let the inclusion probability of the j th psu in the h th stratum be given by

$$p_{PSU} = n_h \cdot \frac{M_{hj}}{\sum_j M_{hj}},$$

and let the inclusion probability of a household be in the j th psu be given by m_{hj}/M_{hj} where

- n_h is the number of psu's in stratum h in the sample;
- M_{hj} is the number of households in the j th psu of the h th stratum;
- $\sum_j M_{hj}$ is the number of households in the h th stratum; and
- m_{hj} is the number of households selected from the j th psu of the h th stratum.

The non-response adjustment factor is given by $1/r_{hj}$, where r_{hj} is the response rate in the j th psu given by

$$r_{hj} = \frac{n_R}{n_h},$$

where n_R is the number of responding households and n_h is the total number of households in the j th psu. Finally, the design weight adjusted for non-response is given by

$$w_{hji} = \frac{1}{n_h} \cdot \frac{\sum_j M_{hj}}{M_{hj}} \cdot \frac{M_{hj}}{m_{hj}} \cdot \frac{n_h}{n_R}$$

where $h = 1, \dots, H$ and $j = 1, \dots, n_h$ [1].

6.4 Simulated Datasets

The IES 2005/2006 survey described in sections 6.1 and 6.2 was used as a surrogate population. Monte Carlo simulation was applied to the surrogate population so that the performance of different weight based estimators of the poverty and inequality measures described in chapter five, could be compared. The bootstrap and jackknife methods were then applied to the simulated data for the purpose of further examining the performance of the estimators, especially through the use of confidence intervals and other measures of accuracy.

The simulation consisted of drawing 1000 samples from the population where each sample has the same design as the IES 2005/2006 survey: a stratified two-stage cluster design with EA's as sampling frame of psu's and the nine provinces as strata. In each selected PSU, six dwelling units were selected.

Differential non-response, for example the under-representation of white people living in urban areas and small households, is found in practical situations in South Africa. Thus to be able to determine this type of non-response error, it was simulated in the design of the samples through the use of auxiliary variables. This was done to evaluate the weighting procedures under non-perfect circumstances. Two sets of auxiliary variables were used in the simulation to aid in determining which weighting technique would be best under these circumstances:

- The first set contains only person level auxiliary variables, indicated by “pp”. These are
 1. Province, with 9 categories;
 2. Gender, with 2 categories; and
 3. Race, with 4 categories.
- The second set contains person and household level auxiliary variables, indicated by “ph”. These are

1. All person level auxiliary variables;
2. Area, with 2 categories; and
3. Household size, with 8 categories.

The application of these simulated datasets as well as any summary measures used to test and compare the accuracy of the different weighting techniques, will be discussed next in chapter seven.

6.5 Conclusion

The Income and Expenditure survey conducted over the period from September 2005 to August 2006 by Statistics South Africa was used as the surrogate population from which smaller datasets were generated by means of Monte Carlo simulation. These simulated datasets will be used in chapter seven to test and compare the accuracy of the different weighting techniques.

Chapter 7

Analysis

7.1 Introduction

In this chapter the Monte Carlo simulation study, as described in section 6.4, is used and the methods described in earlier chapters are applied. Recall that the simulation study consisted of drawing 1000 samples from the population where each sample has the same design as the IES 2005/2006 survey described in chapter 6. Of these samples, the first 100 were considered. It should be noted that all household members within a selected dwelling unit were included in each sample.

There are four types of weighting that will be compared, namely

1. No weighting, thus simple random sampling (None);
2. Household design weight (Design);
3. Raking ratio, integrated weighting method, based on person auxiliary variables (RR_pp); and
4. Raking ratio based on person and household auxiliary variables (RR_ph).

The person level auxiliary variables referred to are

- Province, with 9 categories representing the 9 provinces of South Africa,
- Gender, with categories male and female, and
- Race, with 4 categories,

and the person and household auxiliary variables are

- Province,
- Gender,
- Race,
- Area, with categories urban and rural, and
- Household size, with categories 1, 2, 3, 4, 5, 6, 7 & 8 and 9+.

Let us consider the $R = 100$ samples as 100 bootstrap populations and recall that $\hat{\theta}_r$ is the estimator of the parameter of interest calculated on the r th bootstrap population. In each bootstrap population and for each weighting method, a first level bootstrap was conducted with B , the number of bootstrap samples, equal to 500. These first level bootstrap samples, $\{\mathbf{y}_{br}^*\}$, were used to estimate the variance of the estimators of poverty and inequality, $\hat{V}_B(\hat{\theta}_r)$. For the purpose of constructing the bootstrap- t confidence interval, a second level jackknife was applied to each of the 500 bootstrap samples to estimate the variance of each bootstrap replicate, $\hat{V}_{JK}^*(\hat{\theta}_{br}^*)$. This describes the basic idea underlying the programmes written in R.

The parameters of interest in this study are the poverty and inequality measures for South Africa based on IES 2005/2006. South Africa typically has high poverty rates that should be considered in order to improve upon the methodology currently used. Two poverty lines will be considered, namely

- Absolute poverty line, and
- Relative poverty line.

The most common choice for the absolute poverty line is either the US \$1 a day or US \$2 a day. Here we make use of the US \$1 a day which is approximately equal to a yearly household per-capita income of R2605. The relative poverty line used in this study is the 60% median equivalised income. Recall that equivalised income is calculated as

$$eqinc_i = \frac{totinc_k}{eqsize_k},$$

as defined in equation (5.22). The poverty measures that will be estimated in this study based on the absolute poverty line, are

- Headcount;
- Poverty Gap (Pov Gap); and
- Squared Poverty Gap (Pov Gap2),

while the poverty measures that will be estimated based on the relative poverty line are

- Headcount;
- Poverty Gap (Pov Gap);
- Squared Poverty Gap (Pov Gap2);
- At-risk-of-poverty (AROP), \hat{I}_1 .

The inequality measures that will be estimated based on percy, are

- Gini Coefficient (Gini);
- Mean Log Deviation (MLD); and
- Theil's T,

while those based on eqinc are

- Gini Coefficient (Gini);
- Mean Log Deviation (MLD);
- Theil's T; and
- Quintile Share ratio (QSR), \hat{I}_2 .

The mean log deviation and Theil's T form part of the Generalised Entropy class of inequality measures and \hat{I}_1 and \hat{I}_2 form part of the Laeken indicators. These measures were defined and discussed in chapter 5.

To determine the accuracy of these estimators, certain summary measures were calculated and will be presented. Since we are in the fortunate position of knowing the population, the population value for each of the measures are calculated and denoted by $\hat{\theta}$. It is thus possible to determine the true bias and true mean

squared error of the estimators and these accuracy measures are included in the summary measures. For each of the estimators the following properties are investigated [20][15]:

- The bias of the estimator with respect to the population parameter, θ ,

$$bias(\hat{\theta}) = \left[\left(\frac{1}{R} \sum_{r=1}^R \hat{\theta}_r \right) - \theta \right], \quad (7.1)$$

where R is the number of replicate samples and $\hat{\theta}_r$ is the estimator calculated on the r th replicate sample.

- The mean squared error of the estimator with respect to the population parameter,

$$MSE(\hat{\theta}) = \left[\frac{1}{R} \sum_{r=1}^R (\hat{\theta}_r - \theta)^2 \right]. \quad (7.2)$$

- For each of the R bootstrap populations the bootstrap estimated bias,

$$\widehat{bias}_B(\hat{\theta}_r) = \left(\frac{1}{B} \sum_{b_r=1}^B \hat{\theta}_{b_r}^* \right) - \hat{\theta}_r, \quad (7.3)$$

which results in R bootstrap estimated biases. Thus, the overall bootstrap estimate of bias of $\hat{\theta}$ is given by

$$\widehat{bias}_B(\hat{\theta}) = \left[\frac{1}{R} \sum_{r=1}^R \widehat{bias}_B(\hat{\theta}_r) \right]. \quad (7.4)$$

The difference between the bootstrap estimate of bias and the true bias is then calculated as

$$Dev_{bias}(\hat{\theta}) = \widehat{bias}_B(\hat{\theta}) - bias(\hat{\theta}). \quad (7.5)$$

- Similar to the bootstrap estimate of bias, the bootstrap estimate of MSE is calculated for each of the R bootstrap populations,

$$\widehat{MSE}_B(\hat{\theta}_r) = \frac{1}{B} \sum_{b_r=1}^B \left(\hat{\theta}_{b_r}^* - \hat{\theta}_r \right)^2, \quad (7.6)$$

and the overall bootstrap estimate of MSE of $\hat{\theta}$ is given by

$$\widehat{MSE}_B(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R \widehat{MSE}_B(\hat{\theta}_r). \quad (7.7)$$

The difference between the bootstrap estimate of MSE defined in (7.6) and the true MSE in (7.2) is then calculated as

$$Dev_{MSE}(\hat{\theta}) = \widehat{MSE}_B(\hat{\theta}) - MSE(\hat{\theta}). \quad (7.8)$$

- The percentage relative bias of the estimated MSE of $\hat{\theta}$ with respect to $MSE(\hat{\theta})$,

$$\%RelBias\left(\widehat{MSE}(\hat{\theta})\right) = \left[\frac{\widehat{MSE}_B(\hat{\theta}) - MSE(\hat{\theta})}{MSE(\hat{\theta})} \right] \times 100. \quad (7.9)$$

- The percentage relative bias of the estimated variance of $\hat{\theta}$ with respect to $var(\hat{\theta})$,

$$\%RelBias\left(\widehat{var}(\hat{\theta})\right) = \left[\frac{\frac{1}{R} \sum_{r=1}^R \widehat{var}_B(\hat{\theta}_r) - var(\hat{\theta})}{var(\hat{\theta})} \right] \times 100 \quad (7.10)$$

where

$$\widehat{var}_B(\hat{\theta}_r) = \widehat{MSE}_B(\hat{\theta}_r) - \left[\widehat{bias}_B(\hat{\theta}_r) \right]^2.$$

A substantial part of this study is devoted to the construction of confidence intervals for the estimators of welfare indices. The confidence intervals considered are

1. The standard confidence interval, calculated as

$$\left[\hat{\theta}_r - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{V}_B(\hat{\theta}_r)}{B}}; \hat{\theta}_r + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{V}_B(\hat{\theta}_r)}{B}} \right], \quad (7.11)$$

where $\hat{V}_B(\hat{\theta}_r)$ is the bootstrap estimated variance of $\hat{\theta}_r$ and B is the number of bootstrap replicate samples.

2. The percentile confidence interval, calculated as

$$\left[\hat{\theta}_{r_{lo}}, \hat{\theta}_{r_{up}} \right] = \left[\hat{\theta}_{r_{([B\alpha])}}^*, \hat{\theta}_{r_{([B(1-\alpha)])}}^* \right], \quad (7.12)$$

where $\hat{\theta}_{r_{lo}}$ and $\hat{\theta}_{r_{up}}$ are the $B\alpha$ largest and $B(1-\alpha)$ largest values of the sorted bootstrap replicates, $\{\hat{\theta}_{b_r}^*\}$.

3. The bootstrap- t confidence interval, calculated as

$$\left[\hat{\theta}_r - t_U^* \cdot S_B(\hat{\theta}_r), \hat{\theta}_r - t_L^* \cdot S_B(\hat{\theta}_r) \right], \quad (7.13)$$

where t_U^* and t_L^* are respectively the lower and upper α -points obtained from $t_{(1)}^*, \dots, t_{(B)}^*$, the ordered values of

$$t_{b_r}^* = \frac{(\hat{\theta}_{b_r}^* - \hat{\theta}_r)}{\sqrt{\hat{V}_{JK}(\hat{\theta}_{b_r}^*)}}, \quad (7.14)$$

and $V_{JK}(\hat{\theta}_{b_r}^*)$ is the jackknife estimated variance of $\hat{\theta}_{b_r}^*$.

The theory surrounding these confidence intervals was discussed in chapter 4. The following summary measures were calculated for the different confidence intervals:

- For each of the confidence intervals their non-coverage probability (NCP), measuring the proportion of times that the interval does not contain the true value of the parameter of interest, is calculated. From each of the R bootstrap populations, one confidence interval is calculated for each of standard, percentile and bootstrap- t confidence intervals. This results in 100 standard

confidence intervals, 100 percentile confidence intervals and 100 bootstrap- t confidence intervals. Let $\hat{\theta}_{r_{lo}}$ be the lower limit of the r th confidence interval and let $\hat{\theta}_{r_{up}}$ be the upper limit of the r th confidence interval. This method results in 100 lower limits

$$\hat{\theta}_{1_{lo}}, \hat{\theta}_{2_{lo}}, \dots, \hat{\theta}_{100_{lo}},$$

and 100 upper limits

$$\hat{\theta}_{1_{up}}, \hat{\theta}_{2_{up}}, \dots, \hat{\theta}_{100_{up}},$$

for each of the different confidence intervals methods. Then,

$$\frac{\hat{\theta}_{r_{lo}}}{R} > \hat{\theta}, \quad (7.15)$$

measures the lower non-coverage probability and

$$\frac{\hat{\theta}_{r_{up}}}{R} < \hat{\theta}, \quad (7.16)$$

measures the upper non-coverage probability of each of the different confidence intervals methods.

- The length of the confidence interval is calculated as the difference between the 100 upper limits and lower limits of each different confidence interval method

$$l_r = \hat{\theta}_{r_{up}} - \hat{\theta}_{r_{lo}}, \quad (7.17)$$

resulting in 100 confidence interval lengths

$$l_1, l_2, \dots, l_{100}.$$

for each confidence interval method. The average length of each different confidence interval method,

$$\frac{1}{R} \sum_{r=1}^R l_r,$$

is then plotted for each different weighting method described above to compare the different confidence interval methods. From the length of the confidence intervals their standardized length (Std Lenght) are calculated,

$$\frac{\frac{1}{R} \sum_{r=1}^R l_r}{2z_{\frac{\alpha}{2}} \sqrt{MSE(\hat{\theta})}}. \quad (7.18)$$

- The shape of each of the different confidence intervals is calculated as the distance between each of the 100 upper limits and the corresponding welfare estimator calculated on the r th bootstrap population divided by the distance between the welfare estimator calculated on that r th bootstrap population and each of the 100 lower limits,

$$\hat{\gamma}_r = \frac{\hat{\theta}_{r_{up}} - \hat{\theta}_r}{\hat{\theta}_r - \hat{\theta}_{r_{lo}}}. \quad (7.19)$$

This results in 100 shape estimators for each different confidence interval method,

$$\hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_{100}.$$

The average shape of each different confidence interval method,

$$\frac{1}{R} \sum_{r=1}^R \hat{\gamma}_r,$$

is then plotted for each different weighting method described above to compare the different confidence interval methods.

In section 7.2 the results of the simulation study will be discussed and in section 7.3

the results will be given in tabular form followed by remarks on the tables as well as overall concluding remarks for the welfare indices.

7.2 Discussion of Simulation Results

In this section the results obtained from the Monte Carlo simulation study are examined. All programmes used in the simulation study were written in R and are available from the author at rethak@sun.ac.za. The data is also available at the same address.

7.2.1 Estimators of Welfare Indices

As mentioned before, the population considered here is the Income and Expenditure survey, conducted over the period September 2005 until August 2006. This section considers the estimators and their measures of accuracy. Table 7.1 contains the true values of the welfare indices as calculated from the population and based on both poverty lines. The absolute poverty line, percy, was calculated as

$$percy = \frac{household\ income}{household\ size},$$

and the relative poverty line, equivalized income (Eq Income), was defined in equation (5.22).

WELFARE INDICES	Percy	Eq Income
Headcount	0.29144713	0.27129893
Poverty Gap	0.10816088	0.09830807
Squared Poverty Gap	0.05640399	0.05102522
Gini Coefficient	0.69458962	0.66322250
Mean Log Deviation	0.94146999	0.83142064
Theil T	1.10151793	0.99138040
At-risk-of-poverty	—	0.2712989
Quintile Share Ratio	—	30.4697039

Table 7.1: True values of welfare indices

The at-risk-of-poverty (\hat{I}_1) and quintile share ratio (\hat{I}_2) are only calculated for the relative poverty line. By comparing the values for the different poverty lines it can be seen that there is no substantial difference between them. In table 7.2 the

mean values of the welfare indices, as calculated on each of the 100 replicate samples taken from the population, are given for each weighting method and based on the absolute poverty line, percy. Let $\hat{\theta}_r$ be the estimated welfare index calculated on the r th replicate sample. Then each value in tables 7.2 and 7.3 is calculated as

$$\tilde{\theta} = \frac{1}{R} \sum_{r=1}^R \hat{\theta}_r.$$

WELFARE INDICES	True Value	None	Design	RR_pp	RR_ph
Headcount	0.29144713	0.2902	0.2983	0.2946	0.2912
Poverty Gap	0.10816088	0.1080	0.1114	0.1102	0.1089
Squared Poverty Gap	0.05640399	0.0564	0.0582	0.0576	0.0570
Gini Coefficient	0.69458962	0.6794	0.6797	0.6948	0.6960
Mean Log Deviation	0.94146999	0.8856	0.8856	0.9406	0.9465
Theil T	1.10151793	1.0450	1.0479	1.0963	1.0982

Table 7.2: Mean estimated values of welfare indices under different weighting methods: Percy

The weighted values of each welfare measure are slightly larger than the non-weighted values, but once again no substantial differences are noted. Note that the weighted values generally seem to be quite close to the true values. Table 7.3 contains the mean values based on the relative poverty line, equivalized income, as described in section 7.1:

WELFARE INDICES	True Value	None	Design	RR_pp	RR_ph
Headcount	0.27129893	0.2692	0.2685	0.2785	0.2751
Poverty Gap	0.09830807	0.0987	0.0984	0.1004	0.1014
Squared Poverty Gap	0.05102522	0.0514	0.0512	0.0523	0.0529
Gini Coefficient	0.66322250	0.6478	0.6478	0.6635	0.6633
Mean Log Deviation	0.83142064	0.7826	0.7816	0.8313	0.8312
Theil T	0.99138040	0.9320	0.9332	0.9823	0.9804
At-risk-of-poverty	0.2712989	0.2692	0.2685	0.2785	0.2751
Quintile Share Ratio	30.4697039	27.5876	27.4740	30.4011	30.4959

Table 7.3: Mean estimated values of welfare indices under different weighting methods: Eq Income

Once again the weighted values appear to be slightly larger than the non-weighted values, but quite close to the true values.

Since the true values of the welfare indices are known it is possible to compute the true bias (see eq (7.1)) and true mean square error (as defined in (7.2)) of each index based on each poverty line. These values are given in tables 7.4 and 7.5 and are graphically expressed in figures 7.1 through 7.10.

TRUE BIAS					TRUE MSE				
WELFARE INDICES	None	Design	RR_pp	RR_ph	WELFARE INDICES	None	Design	RR_pp	RR_ph
Headcount	-0.00125	0.00689	0.00319	-0.00029	Headcount	0.00057	0.00065	0.00055	0.00049
Pov Gap	-0.00018	0.0032	0.002	0.00077	Pov Gap	0.00012	0.00014	0.00012	0.00012
Pov Gap2	0.00002	0.00181	0.00123	0.00058	Pov Gap2	0.00005	0.00006	0.00006	0.00005
Gini	-0.01517	-0.01492	0.0002	0.00144	Gini	0.0007	0.00068	0.00054	0.00054
MLD	-0.05584	-0.0559	-0.00089	0.00507	MLD	0.00792	0.00779	0.00624	0.00621
Theil T	-0.05655	-0.05363	-0.00519	-0.00328	Theil T	0.0238	0.02295	0.0269	0.0265

Table 7.4: True Bias and MSE for Welfare Indices under different weighting methods: Percy

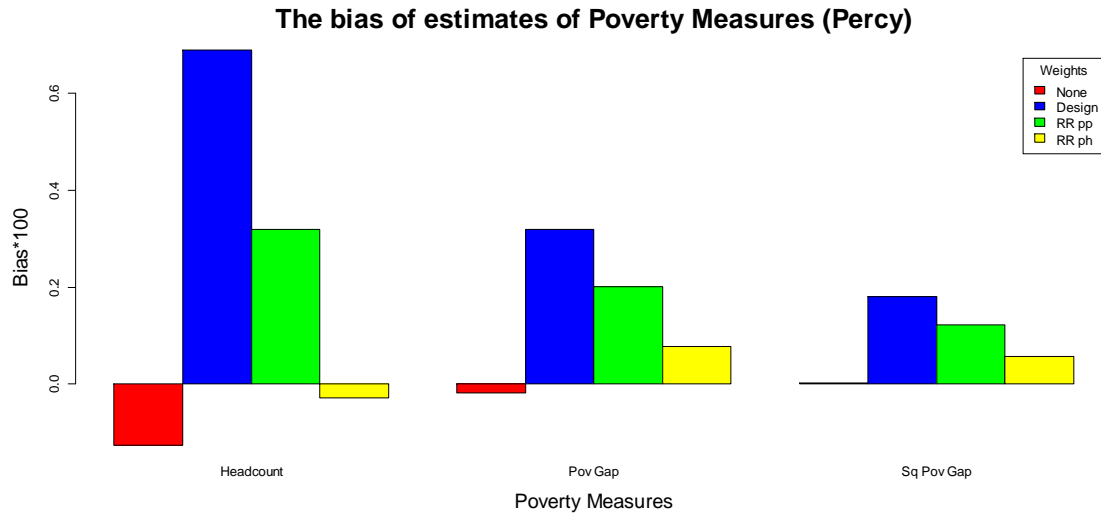


Figure 7.1: True Bias of Estimates of Poverty Measures based on Percy

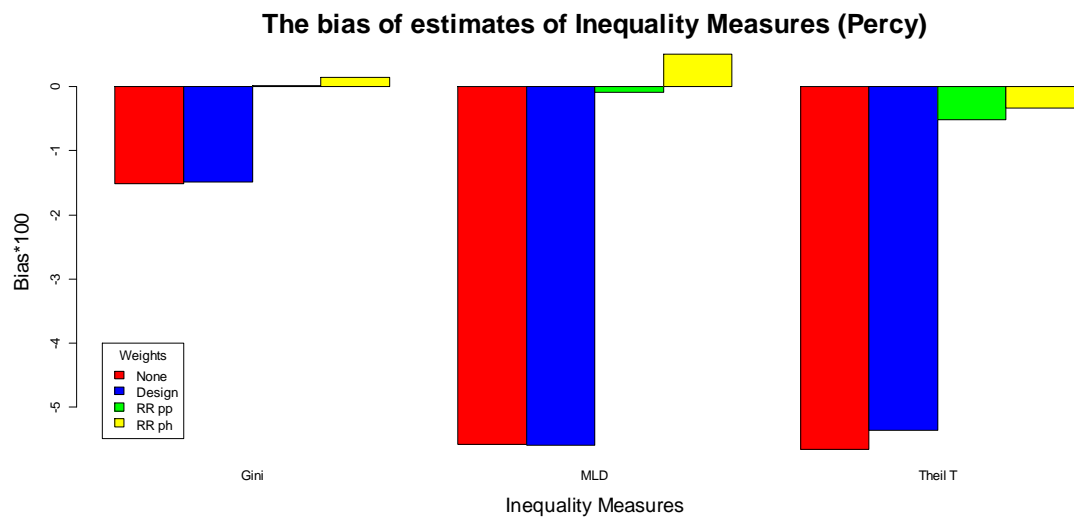


Figure 7.2: True Bias of Estimates of Inequality Measures based on Percy

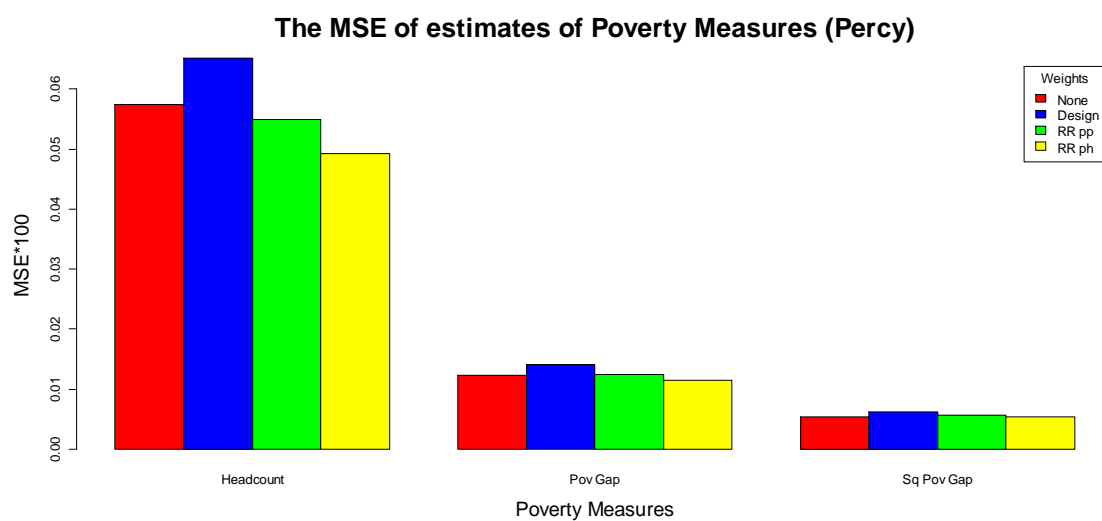


Figure 7.3: True MSE of Estimates of Poverty Measures based on Percy

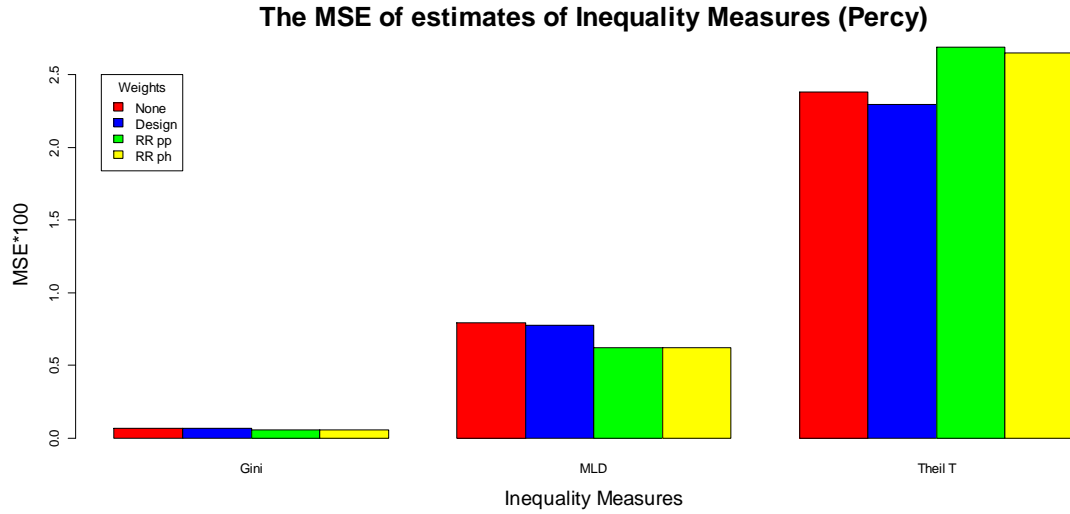


Figure 7.4: True MSE of Estimates of Inequality Measures based on Percy

By studying table 7.4 and figures 7.1 through 7.4 the following conclusions are evident:

- In figure 7.1 it can be seen that the design weight yielded the largest positive bias for each poverty measure while a negative bias was obtained in the case of no weighting.
- Both RR_pp and RR_ph resulted in smaller bias with RR_ph resulting in the smallest bias in figure 7.1.
- In figure 7.2 both no weighting and design weight yielded negative bias for the inequality measures while RR_pp and RR_ph resulted in small bias.
- In figures 7.3 and 7.4 it is clear that weighting resulted in slightly smaller mean squared errors than not weighting, except in the case of the inequality measure, Theil T. This could be contributed to the fact that Theil T is known to be sensitive to outlying values which typically occurred in the population.
- These results coincide with the results in table 7.4.

TRUE BIAS					TRUE MSE				
WELFARE INDICES	None	Design	RR_pp	RR_ph	WELFARE INDICES	None	Design	RR_pp	RR_ph
Headcount	-0.00213	-0.00276	0.00126	0.0038	Headcount	0.00054	0.00055	0.0006	0.00058
Pov Gap	0.00035	0.0001	0.00205	0.00308	Pov Gap	0.00013	0.00013	0.00014	0.00014
Pov Gap2	0.00041	0.00013	0.00127	0.0019	Pov Gap2	0.00005	0.00006	0.00006	0.00006
Gini	-0.01539	-0.01547	0.00031	0.00008	Gini	0.00072	0.00071	0.0006	0.00058
MLD	-0.04883	-0.04981	-0.00014	-0.00021	MLD	0.00645	0.0064	0.00568	0.00548
Theil T	-0.0594	-0.05815	-0.00905	-0.01098	Theil T	0.02355	0.02266	0.02777	0.02647
AROP	-0.00213	-0.00276	0.00126	0.0038	AROP	0.00054	0.00055	0.0006	0.00058
QSR	-2.88214	-2.99568	-0.06865	0.02622	QSR	19.11375	19.01013	16.58666	16.05418

Table 7.5: True Bias and MSE for Welfare Indices under different weighting methods: Eq Income

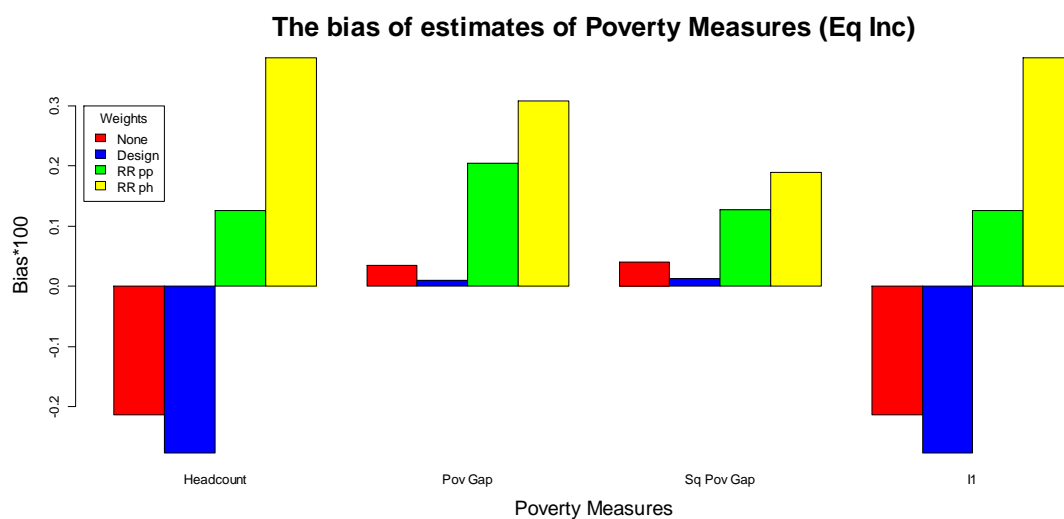


Figure 7.5: True Bias of Estimates of Poverty Measures based on Equivalised Income

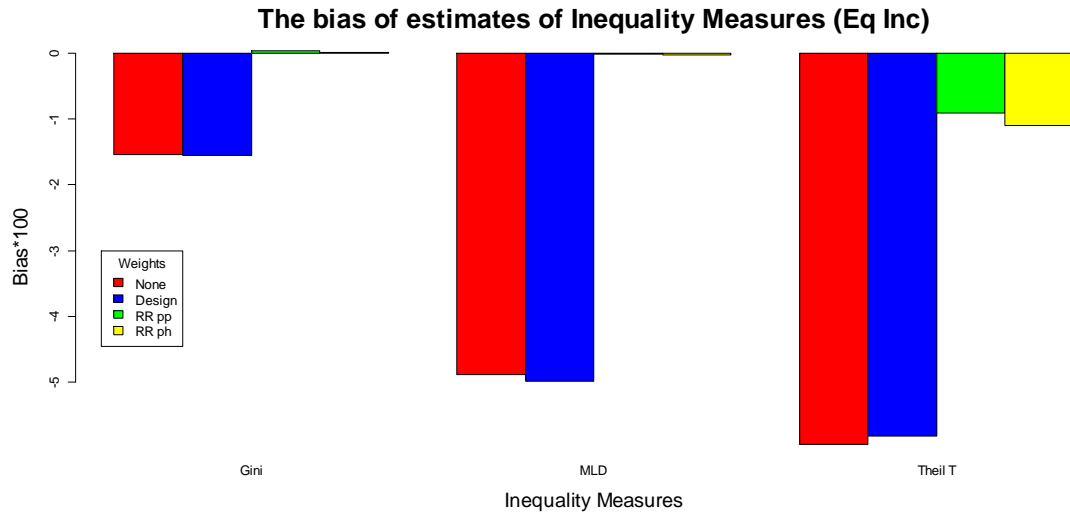


Figure 7.6: True Bias of Estimates of Inequality Measures based on Equivalised Income

Since there seems to be such a large difference between \hat{I}_2 and the other estimated measures due to scale differences, the next figure is only based on \hat{I}_2 . This is done so that the accuracy of the other estimated measures under the different weighting methods can be compared more easily. The procedure is repeated throughout the chapter for all the figures of estimated inequality measures based on the equivalized income poverty line.



Figure 7.7: True Bias of Estimates of Inequality Measures, Equivalised Income: Quintile Share Ratio

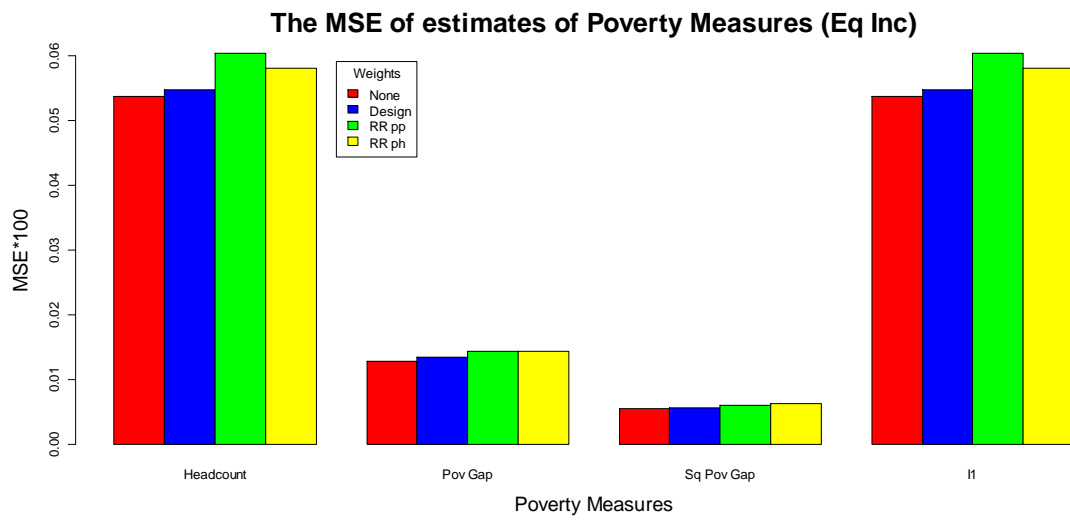


Figure 7.8: True MSE of Estimates of Poverty Measures based on Equivalised Income

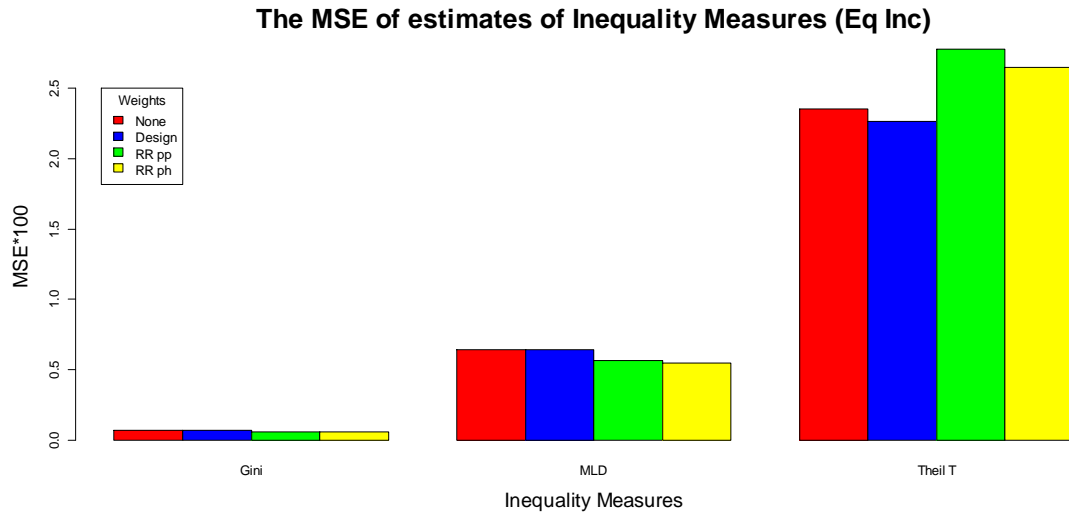


Figure 7.9: True MSE of Estimates of Inequality based on Equivalised Income

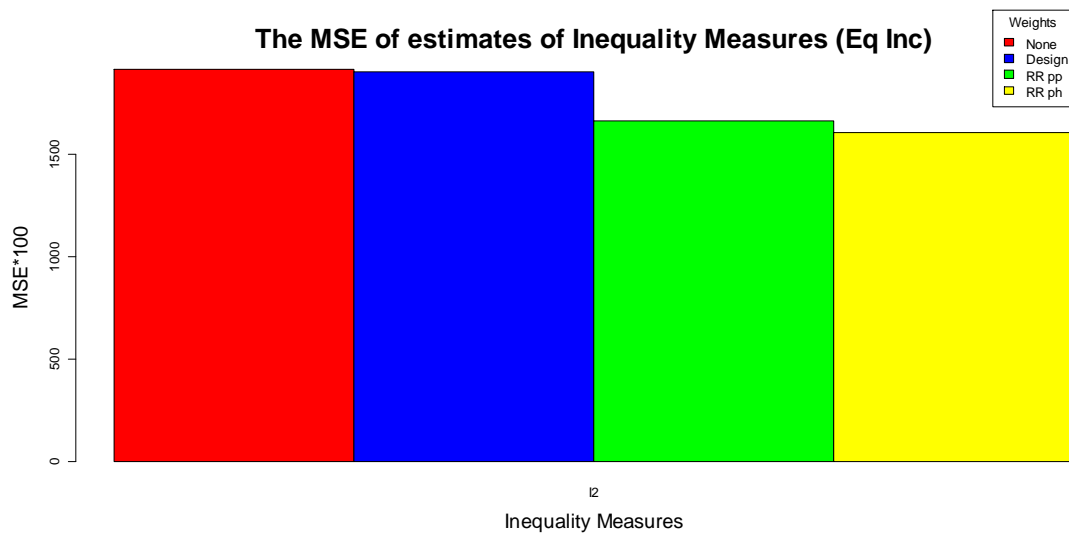


Figure 7.10: True MSE of Estimates of Inequality, Equivalised Income: Quintile Share Ratio

From table 7.5 and figures 7.5 through 7.10 the following conclusions were drawn:

- In figure 7.5 it is concluded that weighting generally resulted in larger bias of poverty measures when these measures were calculated with a relative poverty line. This said, in the case of the headcount index and the at-risk-of-poverty, weighting resulted in over-estimation of the indices which could be regarded

as better than the under-estimation of no weighting. Also, although weighting seemed to give slightly larger bias, these values are still acceptably small.

- In figures 7.6 and 7.7 weighting once again resulted in smaller bias.
- Figure 7.8 shows a slight increase in MSE in comparison to not weighting. Since weighting generally increases variance of estimators this is not an unusual phenomena. Still, these values do appear to be acceptably small and the differences are not really mentionable.
- Figures 7.9 and 7.10 show a slight decrease in MSE for Gini, MLD and quintile share ratio, but once again Theil T shows an opposite pattern.
- These results correspond to the values given in table 7.5.

The true bias and mean squared error can now be compared to the estimated bias and mean squared error obtained using the bootstrap resampling method. Let us recall that the bootstrap method used $B = 500$ resamples from each of the $R = 100$ replicate samples, also called bootstrap populations. The method was described in section 7.1. Tables 7.6 and 7.7 contain the mean estimated bias and mean squared error of each welfare measure calculated for each of the weighting methods. The values in table 7.6 were based on percy while the values in table 7.7 were based on equivalized income.

WELFARE INDICES	None		Design		RR_pp		RR_ph	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Headcount	0.00012	0.00055	0.00012	0.0006	-0.00196	0.00071	-0.00025	0.00066
Poverty Gap	0.00013	0.00011	0.00012	0.00012	-0.00007	0.00015	-0.00006	0.00015
Squared Poverty Gap	0.00008	0.00004	0.00008	0.00005	-0.00034	0.00006	-0.00001	0.00006
Gini Coefficient	-0.00113	0.00047	-0.00113	0.00046	-0.00106	0.00071	-0.00083	0.00069
Mean Log Deviation	-0.00351	0.0048	-0.00358	0.00473	-0.00327	0.00851	-0.00309	0.0084
Theil T	-0.01035	0.01648	-0.01018	0.016	-0.01719	0.02607	-0.01539	0.0255

Table 7.6: Mean Estimated Bias and MSE for Welfare Indices under different weighting methods with the bootstrap resampling method: Percy

Note that these are the only welfare estimators calculated on a absolute poverty line and hence only their results are quoted in this table. By studying table 7.6 the following conclusions can be drawn:

- Although not a consistent trend, the mean estimated bias tends to be slightly smaller for some of the measures under the weighting methods compared to when no weighting is done.
- The mean estimated mean squared error is only slightly larger under the weighting methods than for no weighting.
- For both bias and mean squared error, no weight and design weight show similar results while RR_pp and RR_ph show similar results.

Along with table 7.6 graphs were made of the difference between the bootstrap estimated bias/MSE and the true bias/MSE for the absolute poverty line and are given in figures 7.11 through 7.14. Refer to section 7.1 for a description of how the values of these graphs were obtained. These graphs lend a more intuitive approach to the accuracy of the estimators under each weighting method.

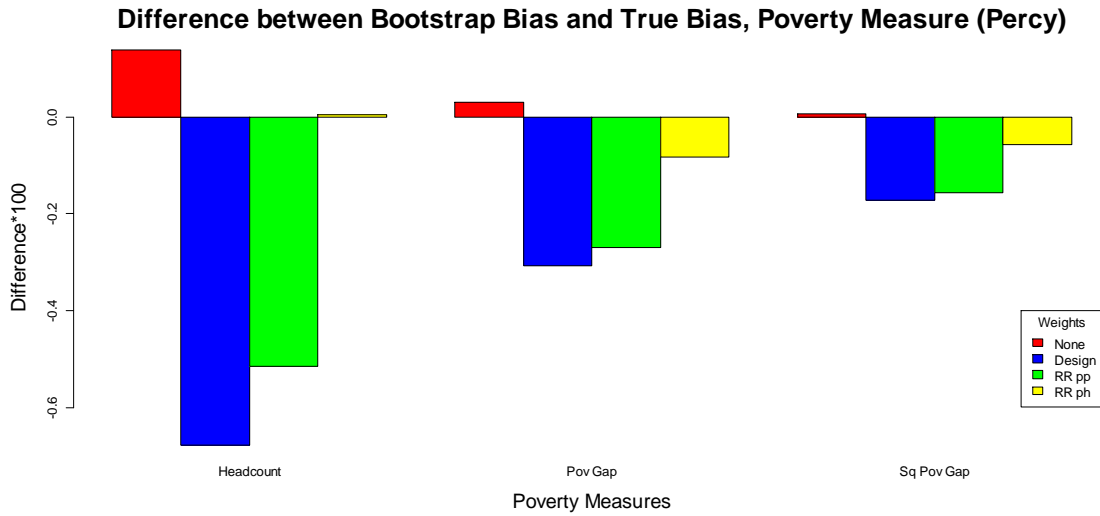


Figure 7.11: Difference between Bootstrap Bias and True Bias of Poverty Measures (Percy)

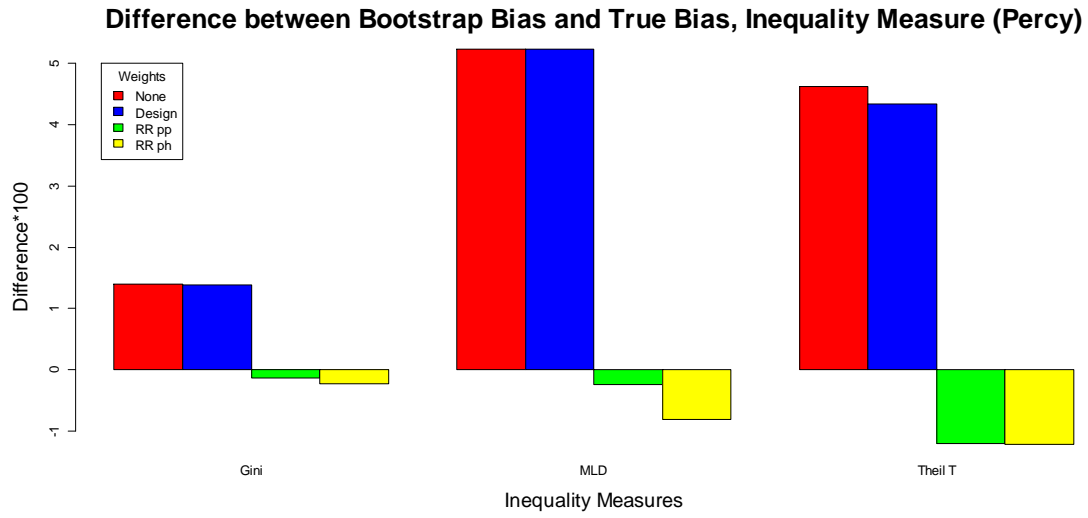


Figure 7.12: Difference between Bootstrap Bias and True Bias of Inequality Measures (Percy)

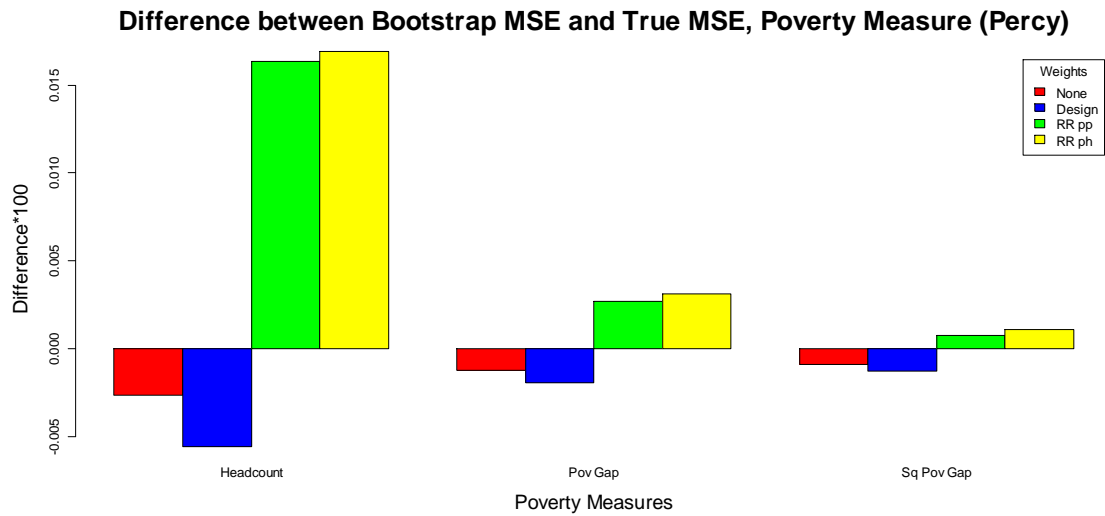


Figure 7.13: Difference between Bootstrap MSE and True MSE of Poverty Measures (Percy)

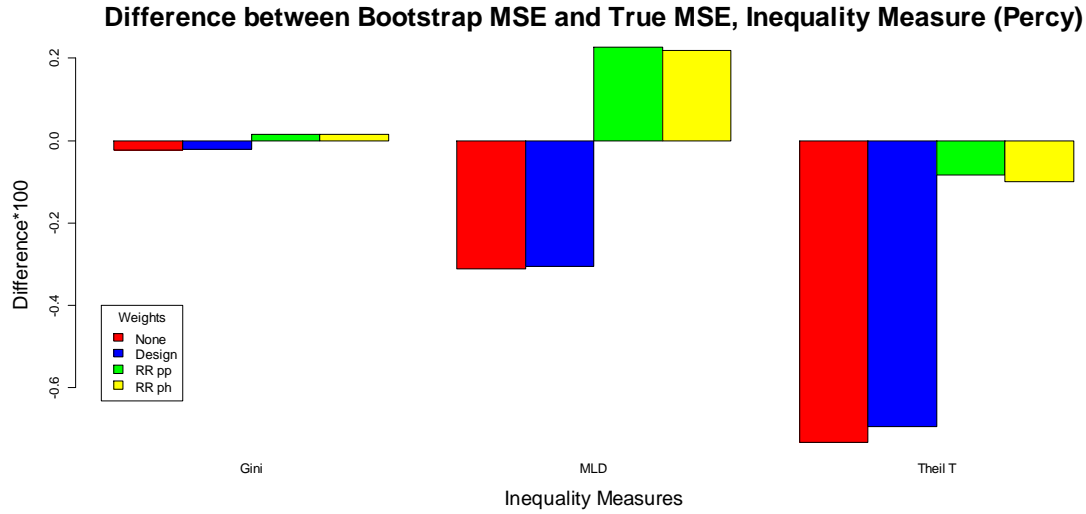


Figure 7.14: Difference between Bootstrap MSE and True MSE of Inequality Measures (Percy)

The following can be concluded from figures 7.11 through 7.14:

- Figure 7.11 shows a decrease in the difference between the bootstrap estimated bias and the true bias of the poverty measures over the different weighting methods. No weight resulted in a smaller difference than the other different weighting methods.
- In figure 7.12 it is seen that no weight and design weight yield similar differences while RR_pp and RR_ph show similar differences. There is also a decrease in the differences to the extent of negative differences under RR_pp and RR_ph. Thus, the mean bootstrap estimated bias under these two weighting methods was smaller than the true bias under these weighting methods.
- Figure 7.13 shows that the difference between the mean bootstrap estimated MSE and the true MSE became larger as the weighting methods were applied. Once again this is expected since weighting usually results in slightly larger variances. Except for Headcount, the other differences were quite small even with the slight increases that occurred.
- In figure 7.14, for the inequality measures, the difference decreased slightly as the weighting methods were applied. For Gini and MLD there was not much difference, but Theil T showed improvement under RR_pp and RR_ph.

Table 7.7 shows the results of table 7.6 based on the relative poverty line.

WELFARE INDICES	None		Design		RR_pp		RR_ph	
	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
Headcount	0.00069	0.00072	0.0018	0.00076	0.00074	0.00124	-0.00111	0.00126
Poverty Gap	0.00056	0.00014	0.00061	0.00015	0.00032	0.00027	-0.0002	0.00027
Squared Poverty Gap	0.00033	0.00006	0.00036	0.00006	0.00025	0.00012	-0.00002	0.00013
Gini Coefficient	-0.00334	0.00075	-0.00339	0.00074	-0.00776	0.0015	-0.00723	0.00146
Mean Log Deviation	-0.00482	0.00643	-0.005	0.00633	-0.01297	0.0154	-0.01175	0.01493
Theil T	-0.01443	0.02106	-0.01424	0.02053	-0.03477	0.03968	-0.03275	0.03844
At-risk-of poverty	-0.00012	0.0006	0.00064	0.00062	-0.00074	0.00135	-0.00193	0.00134
Quintile Share Ratio	-0.30189	18.65189	-0.37306	18.39575	-0.93357	60.30619	-0.87068	58.21672

Table 7.7: Mean Estimated Bias and MSE for Welfare Indices under different weighting methods with the bootstrap resampling method: Eq Income

By studying table 7.7 the following conclusions can be drawn:

- When comparing the mean bootstrap estimated bias of each welfare measure under each weighting method it is seen that there is a slight decrease for certain measures in these values. No weight and design weight seemed to give smaller biases than RR_pp and RR_ph.
- The mean bootstrap estimated MSE of each welfare measure tends to increase slightly over the different weighting methods, as expected.

Figures 7.15 through 7.17 show graphically the difference between the bootstrap estimated bias and true bias based on the relative poverty line, followed in figures 7.18 through 7.20 by the differences between the estimated bootstrap MSE and true MSE.

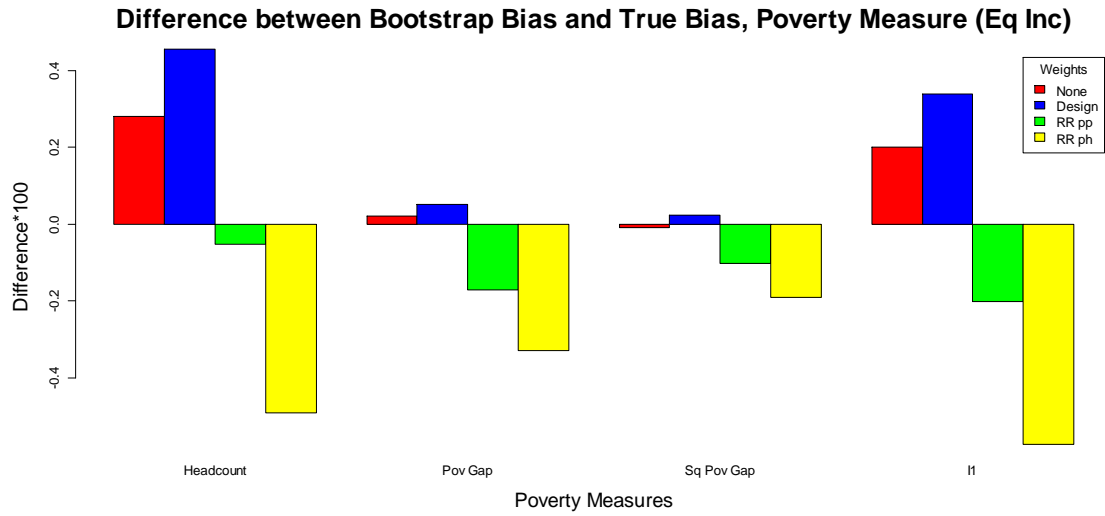


Figure 7.15: Difference between Bootstrap Bias and True Bias of Poverty Measures (Equivalised Income)

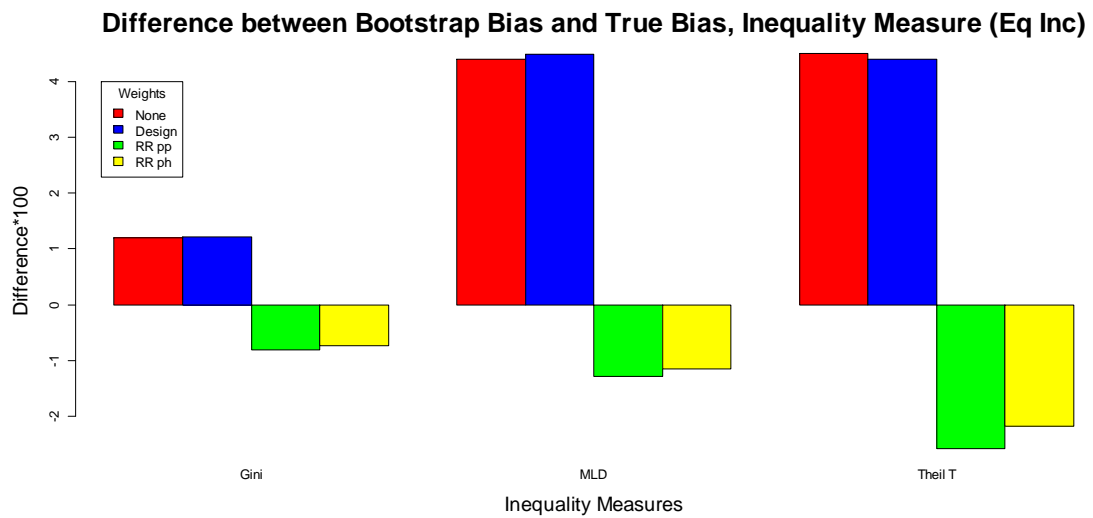


Figure 7.16: Difference between Bootstrap Bias and True Bias of Inequality Measures (Equivalised Income)

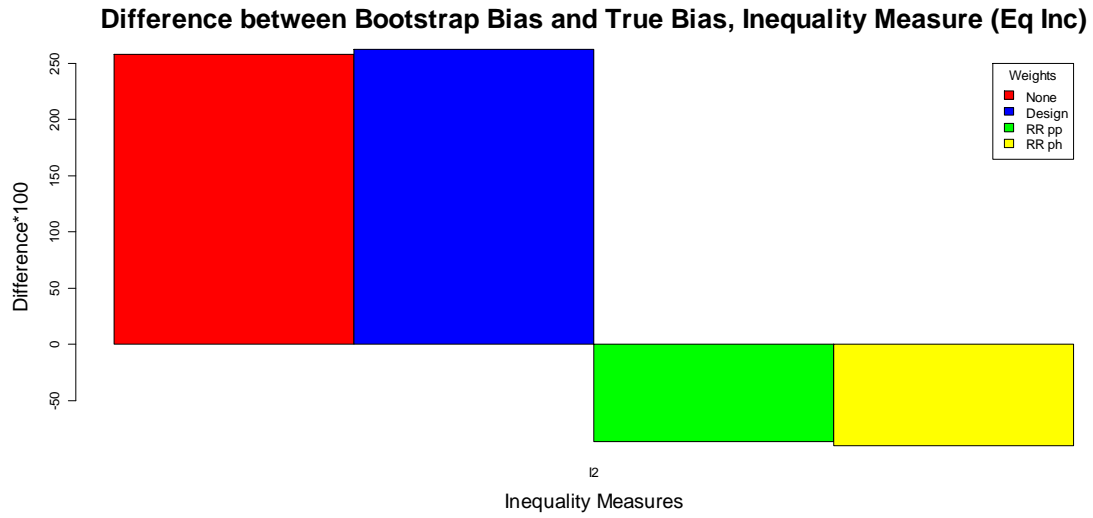


Figure 7.17: Difference between Bootstrap Bias and True Bias of Inequality Measures (Equivalised Income): Quintile Share Ratio

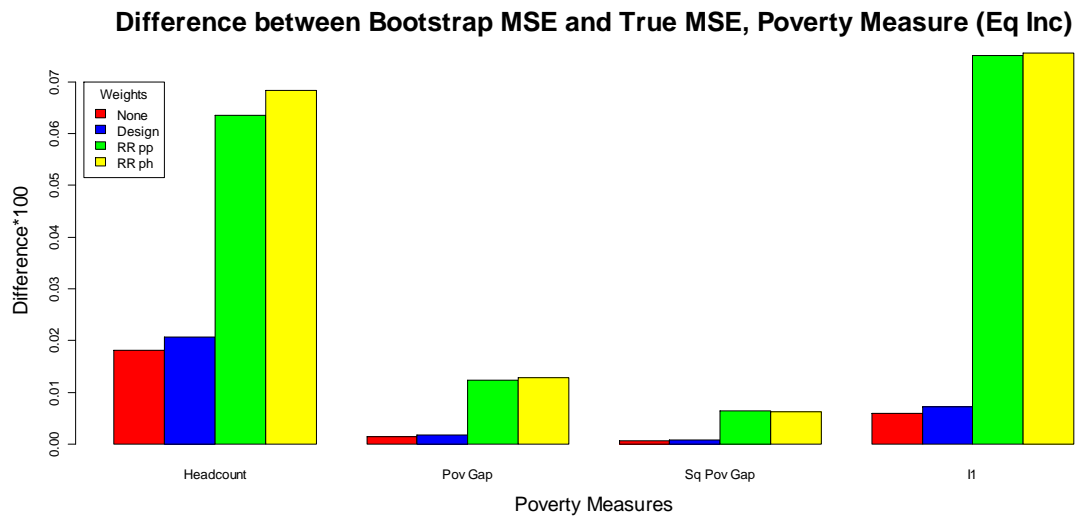


Figure 7.18: Difference between Bootstrap MSE and True MSE of Poverty Measures (Equivalised Income)

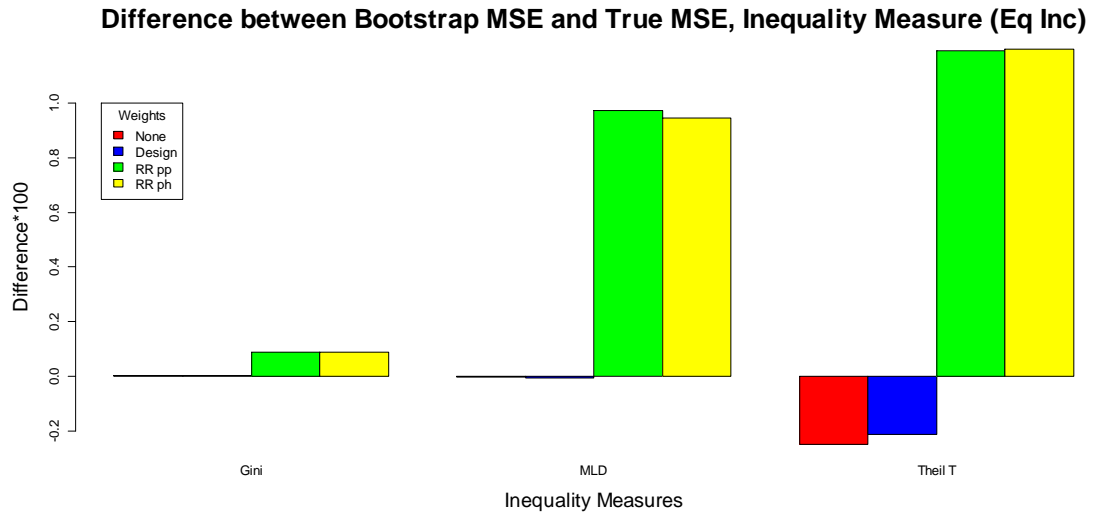


Figure 7.19: Difference between Bootstrap MSE and True MSE of Inequality Measures (Equalised Income)

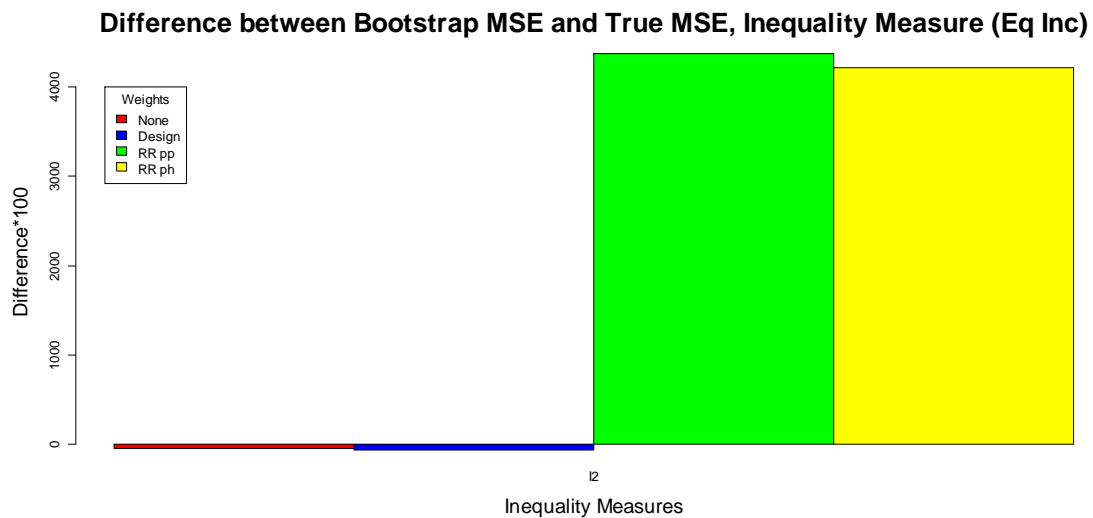


Figure 7.20: Difference between Bootstrap MSE and True MSE of Inequality Measures (Equalised Income): Quintile Share Ratio

The following can be concluded from the figures:

- All figures show that no weight and design weight are similar while RR_pp and RR_ph are similar, except fig. 7.15.

- In figures 7.15, 7.16 and 7.17 no weight and design weight resulted in positive differences while RR_pp and RR_ph resulted in slightly negative differences. Although negative, RR_pp resulted in smaller absolute differences than no weight and design weight while RR_ph gave opposite results.
- Figures 7.18 through 7.20 show a general increase in the difference between the bootstrap estimated MSE and the true MSE of the inequality measures over the different weighting methods. Design weight shows very similar results to no weight while RR_pp tends to be slightly smaller than RR_ph.

The final measures of accuracy to consider in this section are those of the percentage relative bias of a variance estimator as well as the percentage relative bias of a mean squared error estimator. Refer to equations (7.9) and (7.10) respectively.

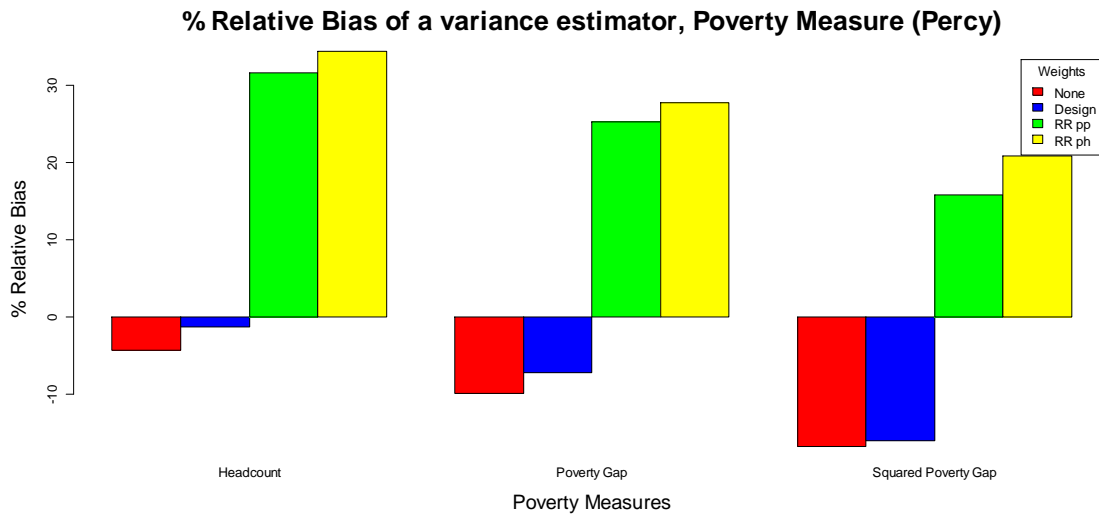


Figure 7.21: % Relative Bias of a variance estimator of Poverty Measures (Percy)

By studying figure 7.21 the following conclusions can be made:

- RR_pp and RR_ph yielded the largest percentage relative bias, values above 25%. RR_pp yielded slightly better relative bias than RR_ph.
- The design weight alone produced less biased estimators with percentage relative bias of less than -10%.
- None of these percentages are however negligible.

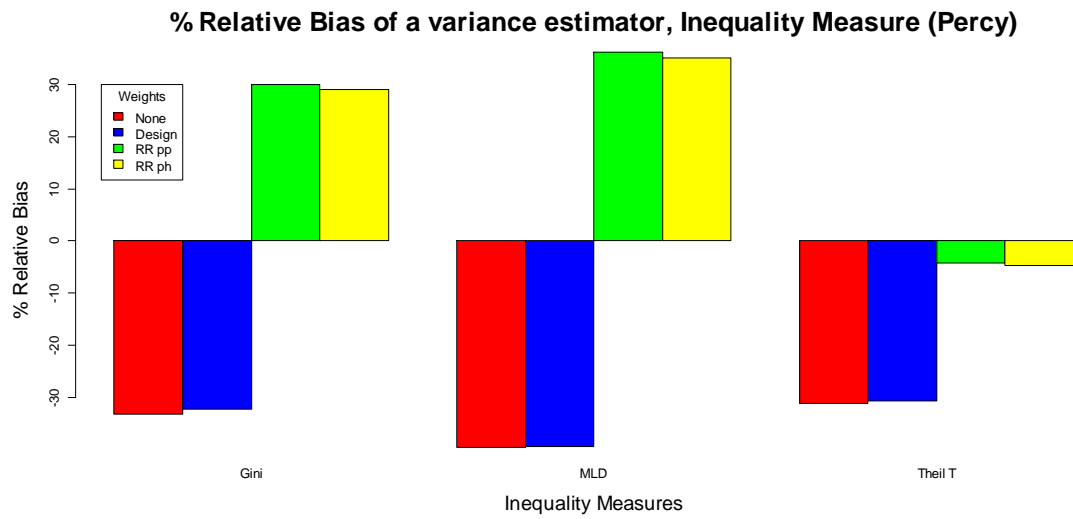


Figure 7.22: % Relative Bias of a variance estimator of Inequality Measures (Percy)

The following conclusions can be drawn from figure 7.22:

- The estimation of the inequality measures based on the absolute poverty line yielded somewhat smaller biases.
- It would seem as if RR_pp and RR_ph produced estimators with the smallest absolute bias, especially in the estimation of Theil T with percentage bias of less than -10%.
- Still, the percentages are not negligible.

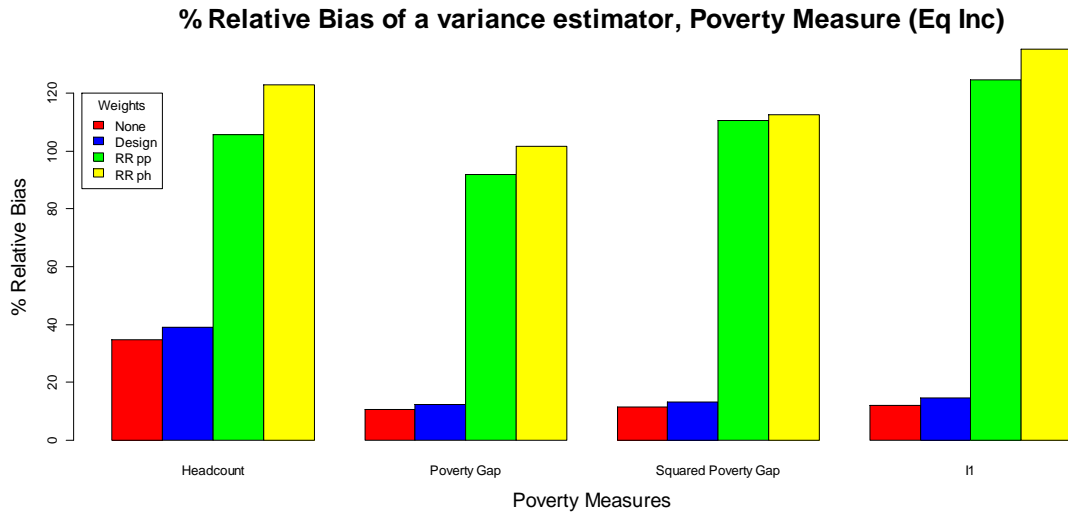


Figure 7.23: % Relative Bias of a variance estimator of Poverty Measures (Eq Income)

Conclusions that can be drawn from figure 7.23:

- Estimation of welfare measures with RR_pp and RR_ph weighting, based on the relative poverty line, yielded very high relative bias with respect to their true variances. Some of the values are above 120%.
- Design weight produced less biased estimators with percentage relative bias of less than 20%, but slightly larger than no weight.
- Again these percentages are not negligible.

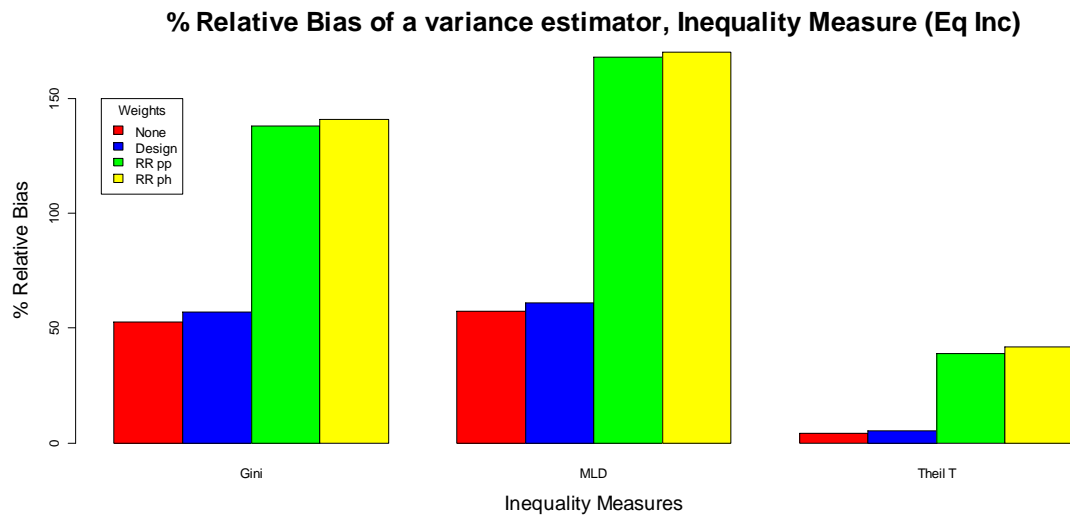


Figure 7.24: % Relative Bias of a variance estimator of Inequality Measures (Eq Income)

Figure 7.24:

- Similar results to figure 7.23, but here RR_ph produced bias estimators with relative bias of greater than 150%.
- RR_pp produced less biased estimators, but only slightly smaller than RR_ph.
- Design weight produced estimators with less bias, but slightly larger than no weight.

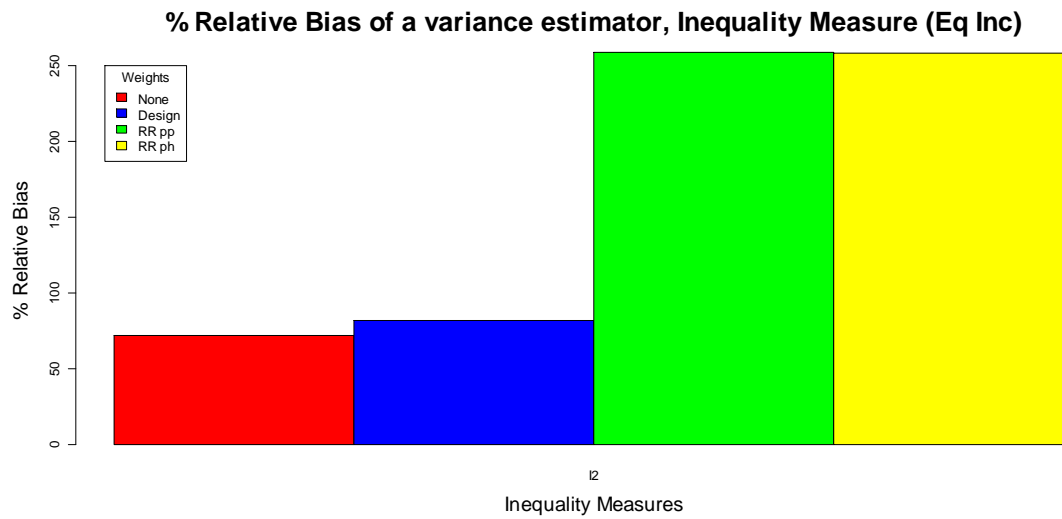


Figure 7.25: % Relative Bias of a variance estimator of Poverty Measures (Eq Income): Quintile Share Ratio

By examining figure 7.25 the following conclusions can be drawn:

- RR_pp and RR_ph, calibrated weighting techniques based on person and person and household variables, produced estimators with very high relative bias with values above 250%.
- Design weight produced estimators with less bias than the calibrated weighting techniques, but slightly larger than no weight.

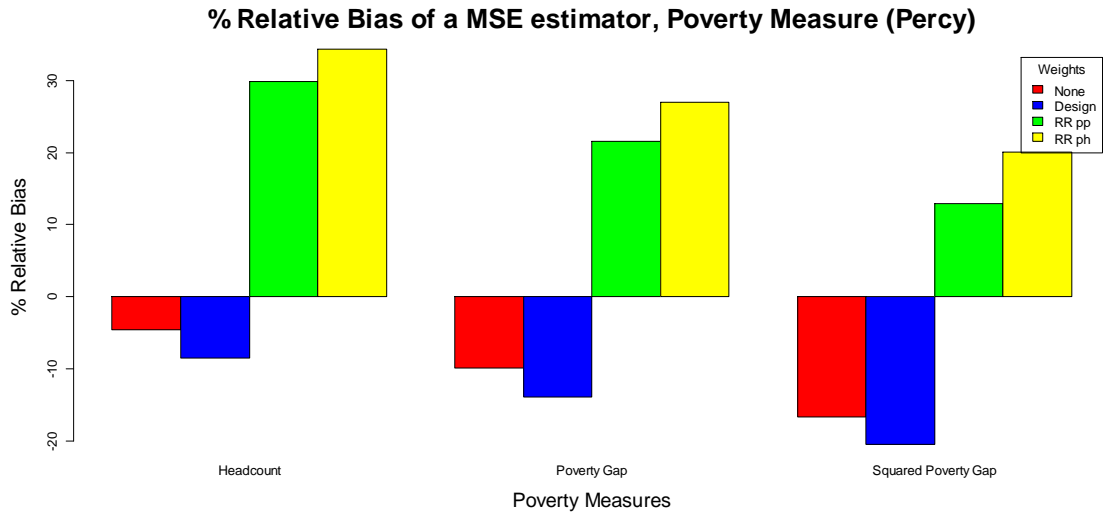


Figure 7.26: % Relative Bias of a MSE estimator of Poverty measures (Percy)

In figure 7.26 it can be seen that RR_pp and RR_ph produced the largest bias with RR_ph producing a relative bias of larger than 30%, except for squared poverty gap.

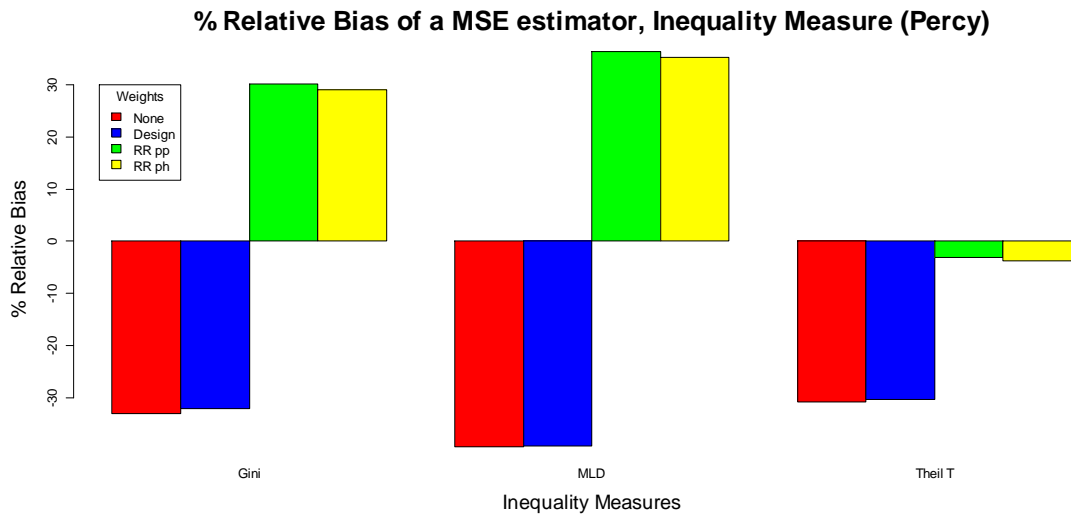


Figure 7.27: % Relative Bias of a MSE estimator of Inequality measures (Percy)

The conclusions that can be drawn from figure 7.27 are that RR_pp and RR_ph appeared to produce the smallest bias, but still some relative biases exceeded 30%. Improvement is shown from no weighting.

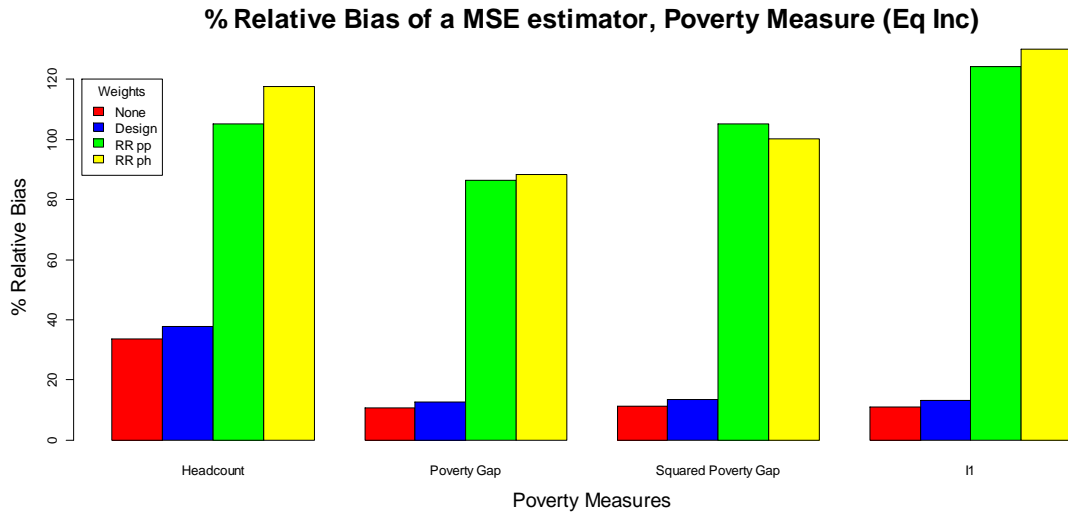


Figure 7.28: % Relative Bias of a MSE estimator of Poverty measures (Eq Income)

In figure 7.28 it is seen that design weight performed better than RR_pp and RR_ph, but slightly larger than no weight. None of its relative bias percentages exceed 20%.

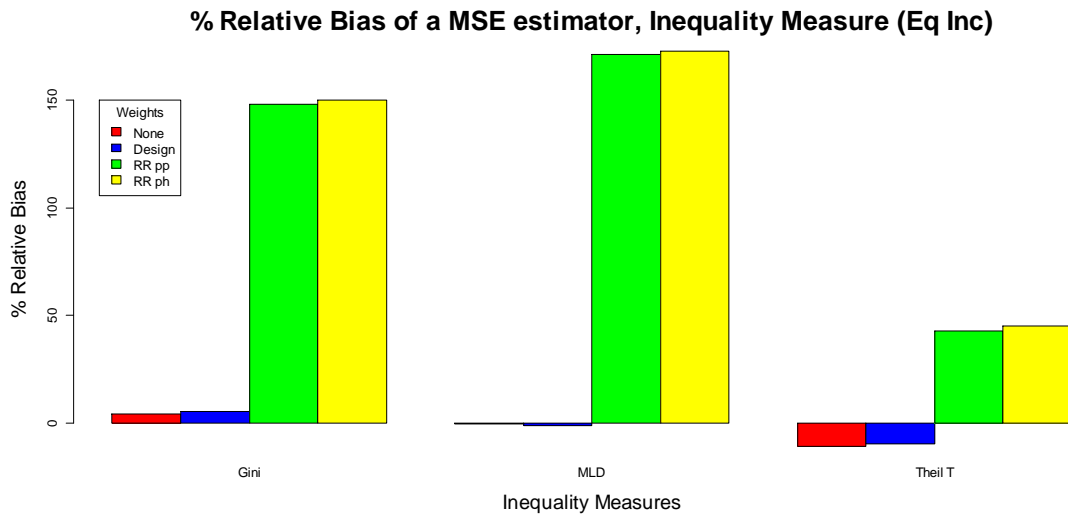


Figure 7.29: % Relative Bias of a MSE estimator of Inequality measures (Eq Income)

Examining figure 7.29 shows similar results to figure 7.28. Design weight produces the smallest bias, but RR_pp and RR_ph exceed 150% in the case of the MLD. No weight and design weight induced much smaller percentage relative bias

than RR_pp and RR_ph with design weight only slightly larger than no weight.

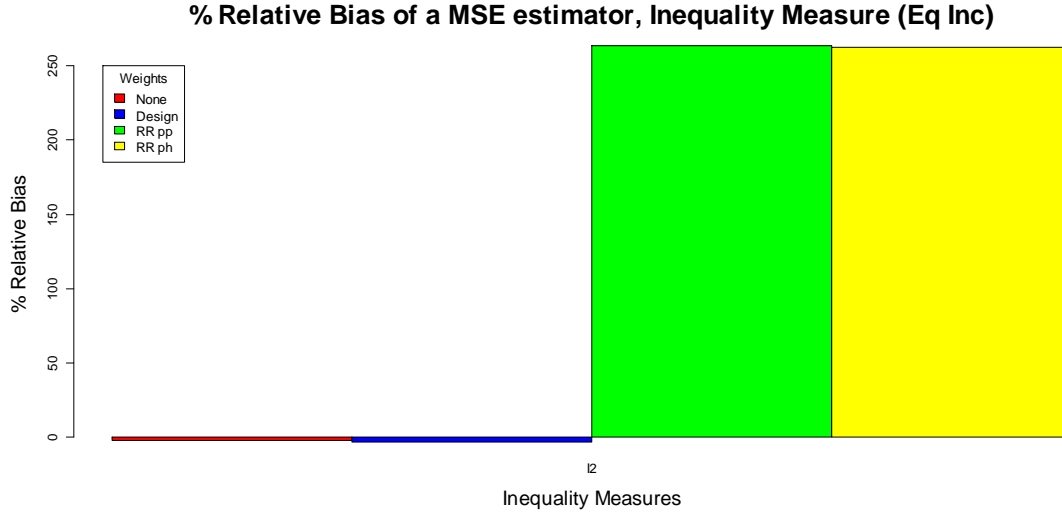


Figure 7.30: % Relative Bias of a MSE estimator of Inequality measures (Eq Income), without Quintile Share Ratio

From figure 7.30 it can be concluded that, of the weighting procedures, design weight produced the smallest bias, while RR_pp and RR_ph once again produced high relative bias percentages. No weight and design weight has much smaller percentage relative bias and design weight is only slightly larger than no weight.

7.2.2 Confidence Intervals for Estimators of Welfare Indices

This section reviews the results obtained for the 95% confidence intervals constructed for the estimators of the welfare indices, along with their measures of accuracy. Recall that the confidence intervals considered are:

- the standard (asymptotic) confidence interval, equation (7.11);
- the percentile confidence interval, equation (7.12); and
- the bootstrap- t confidence interval, equation (7.13).

The following notation should be introduced for some of the graphs presented in this section:

- red dotted line represents the estimated welfare index based on a particular bootstrap population;

- purple lines represent the standard (asymptotic) confidence interval;
- blue lines represent the percentile confidence interval; and
- green lines represent the bootstrap- t confidence interval.

Figures 7.31 through 7.38, are included to illustrate the difference in coverage of these confidence intervals. The figures were calculated for the 31st bootstrap population since it was the sample used at the time as a “test” sample. The histograms in these figures represent the distributions of the 500 bootstrap estimators of each welfare measure. Note the following in the figures:

- The degree of skewness typically found in the case of the relative poverty line;
- The standard (asymptotic) confidence interval is very conservative;
- The percentile interval is wider than the standard confidence interval, but more or less centered with regard to the estimated parameter of interest;
- The bootstrap- t confidence interval is generally skew with respect to the estimated parameter of interest.

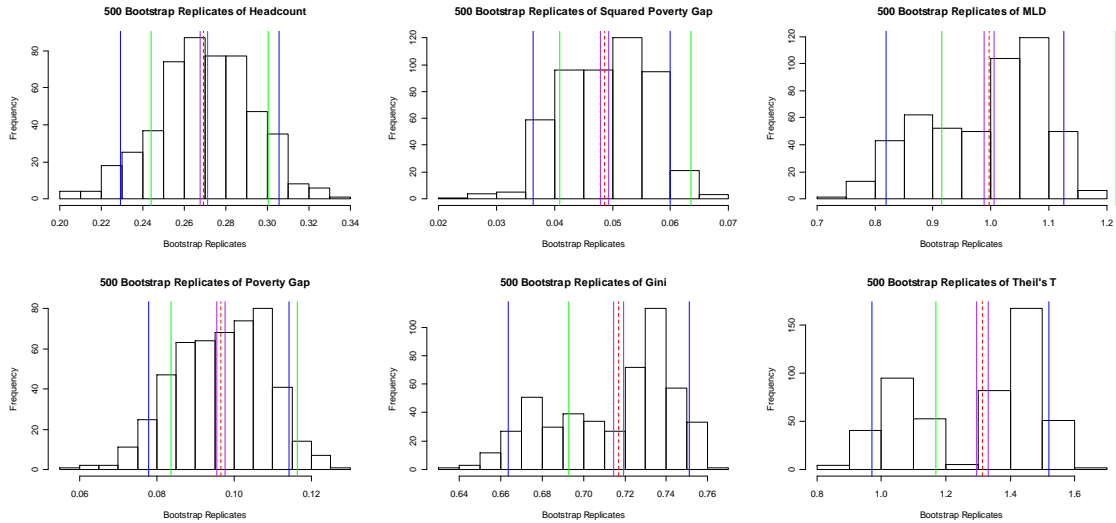


Figure 7.31: Comparison of 95% Confidence Intervals: No Weighting (Percy)

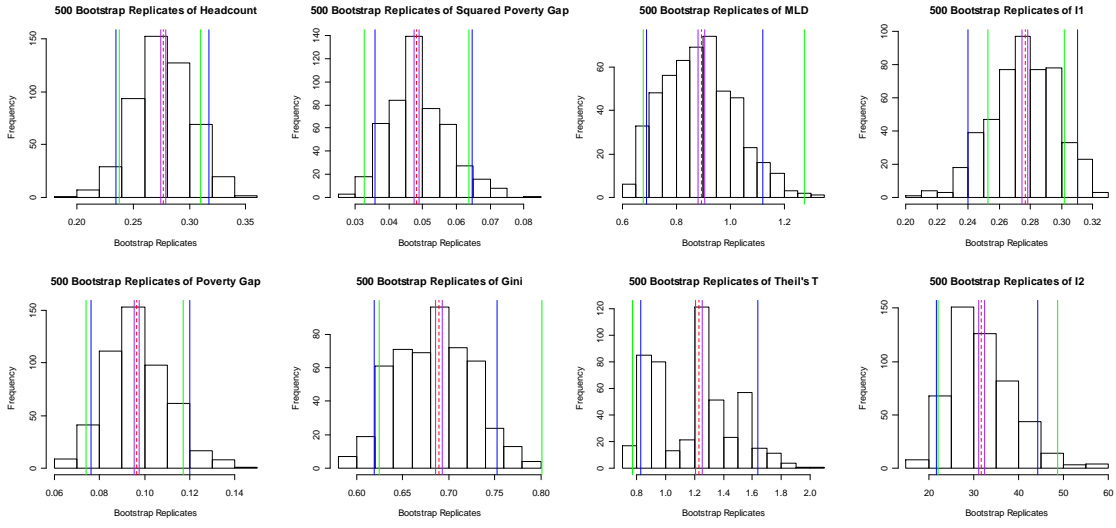


Figure 7.32: Comparison of 95% Confidence Intervals: No Weighting (Eq Income)

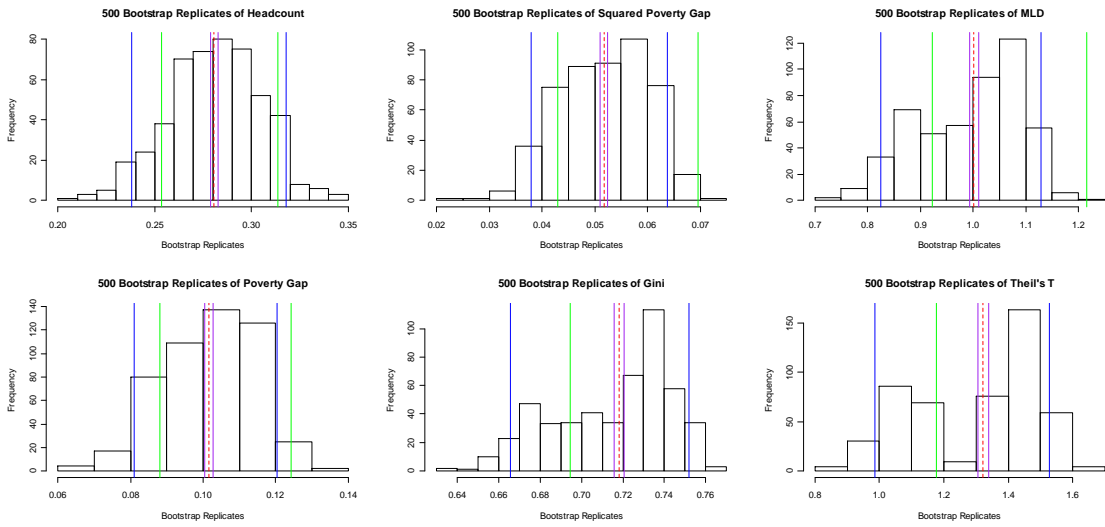


Figure 7.33: Comparison of 95% Confidence Intervals: Design Weight (Percy)

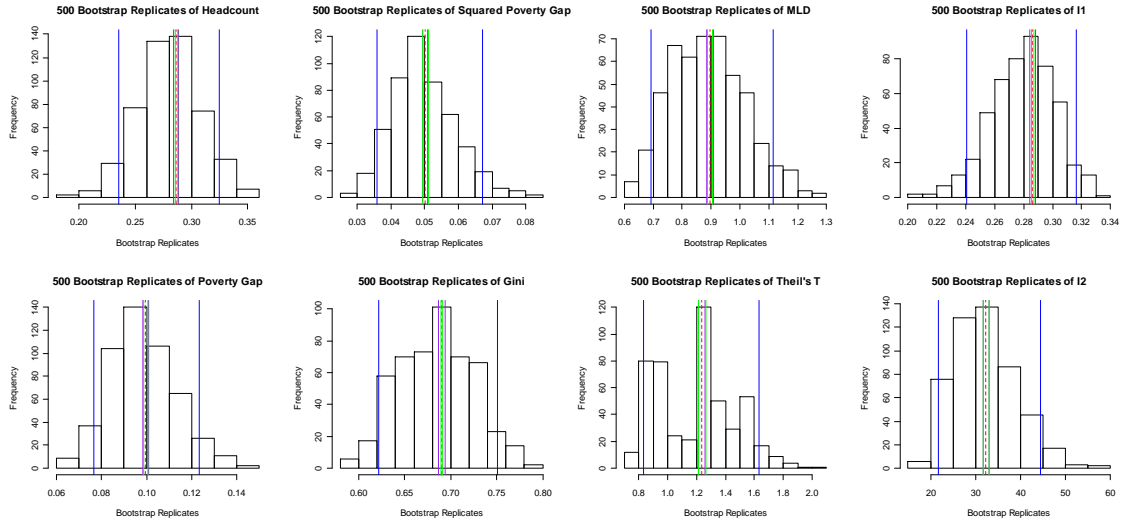


Figure 7.34: Comparison of 95% Confidence Intervals: Design Weight (Eq Income)

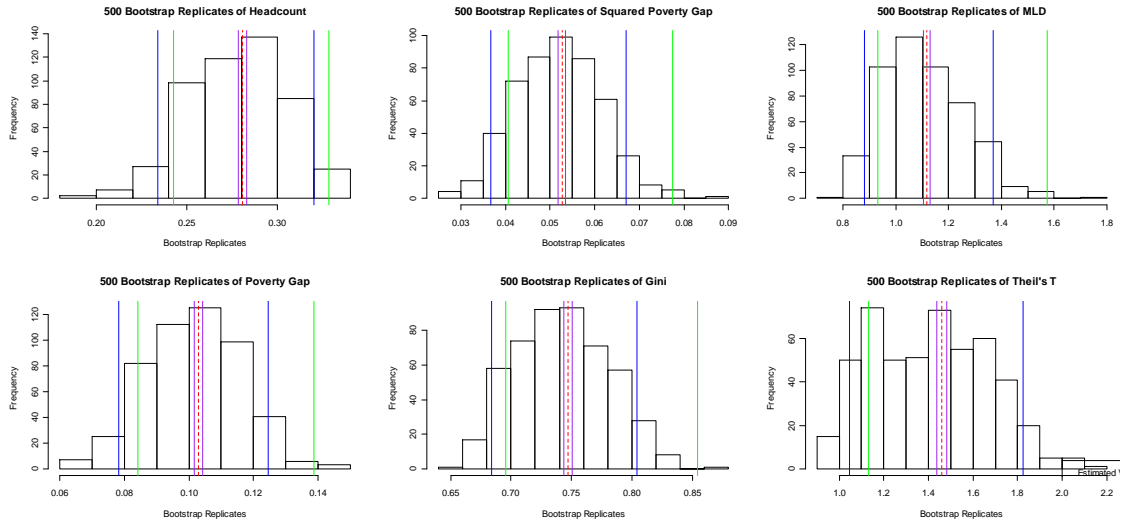


Figure 7.35: Comparison of 95% Confidence Intervals: RR_pp (Percy)

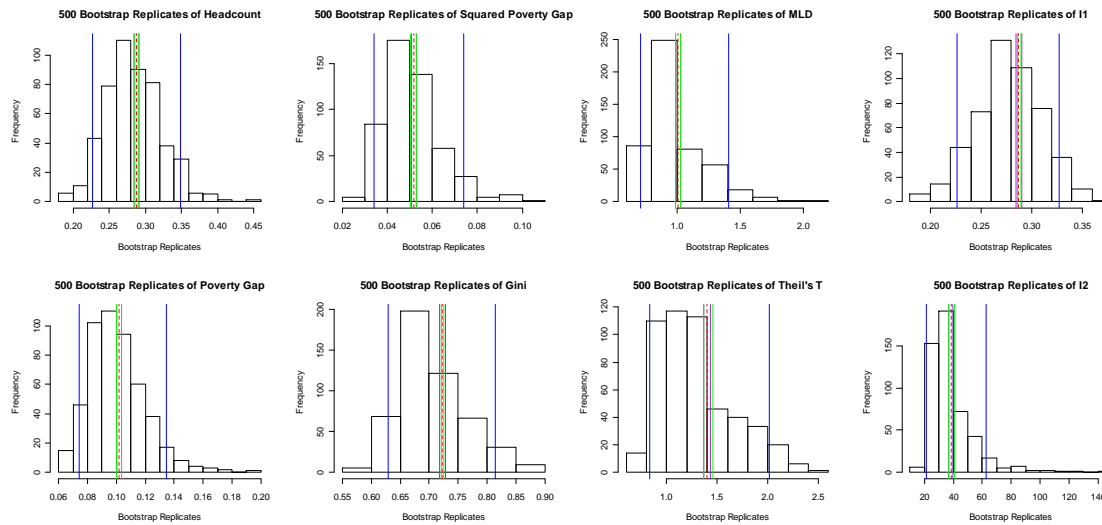


Figure 7.36: Comparison of 95% Confidence Intervals: RR_{pp} (Eq Income)

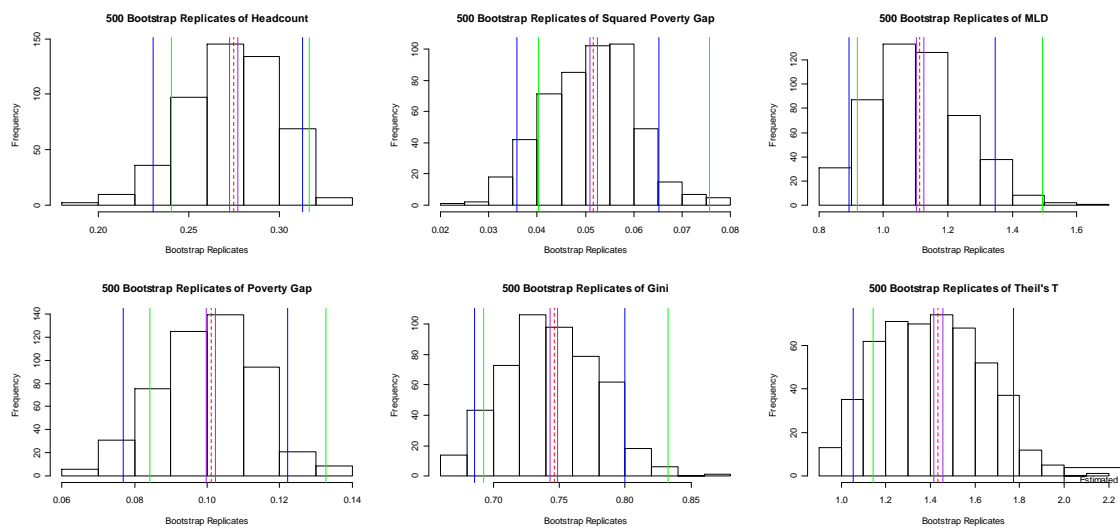


Figure 7.37: Comparison of 95% Confidence Intervals: RR_{ph} (Percy)

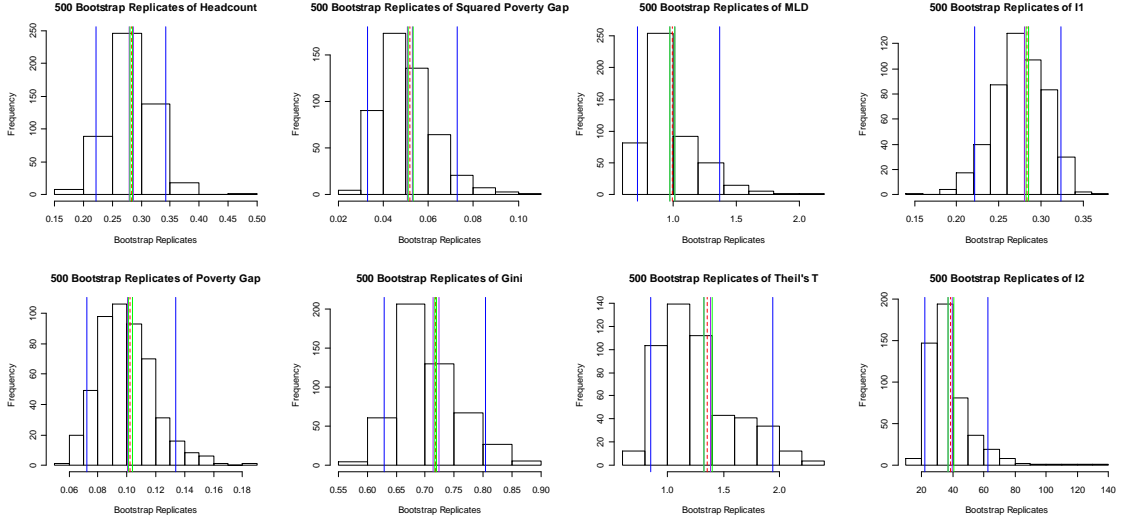


Figure 7.38: Comparison of 95% Confidence Intervals: RR_ph (Eq Income)

The most important assessment of the accuracy of confidence intervals is to consider their coverage probability. Here we measured the non-coverage probability, otherwise known as the complement of the coverage probability, of each interval with regard to the tails of the intervals. The lower non-coverage probability is the proportion of times that the lower limit of the confidence interval is larger than the true value of the parameter of interest. The upper non-coverage probability is the proportion of times that the upper limit of the confidence interval is smaller than the true value of the parameter of interest. Refer to equations (7.15) and (7.16) in section 7.1. Thus, in a nutshell, it measures the proportion of times that the confidence interval does not contain the true value of the parameter of interest.

Table. 7.8 contains the non-coverage probabilities of the standard (asymptotic) confidence interval. Both poverty lines' results are included in the table. The values omitted correspond to the Laeken indices that are not measured with regard to the absolute poverty line.

	None		Design		RR_pp		RR_ph	
WELFARE INDICES	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Headcount: Percy	0.4330	0.4742	0.5773	0.3505	0.5258	0.3918	0.4845	0.4536
Headcount: Eq Income	0.4742	0.4227	0.4639	0.4742	0.5464	0.3505	0.6186	0.2990
Pov Gap: Percy	0.4536	0.4948	0.5361	0.3814	0.5155	0.4124	0.4948	0.4227
Pov Gap: Eq Income	0.4742	0.4536	0.4536	0.4536	0.5155	0.4021	0.5155	0.4277
Pov Gap2: Percy	0.4433	0.5052	0.5052	0.4277	0.4845	0.4227	0.4536	0.4845
Pov Gap2: Eq Income	0.4948	0.4330	0.4742	0.4639	0.5052	0.4124	0.5258	0.3814
Gini: Percy	0.2165	0.7113	0.2268	0.7216	0.4433	0.4742	0.4845	0.4536
Gini: Eq Income	0.1959	0.7629	0.1649	0.7732	0.4021	0.4948	0.4124	0.4639
MLD: Percy	0.1959	0.7526	0.1856	0.7423	0.3814	0.5052	0.4536	0.4536
MLD: Eq Income	0.1856	0.7732	0.1649	0.7835	0.3711	0.5052	0.3918	0.4742
Theil T: Percy	0.2990	0.6804	0.3093	0.6701	0.4124	0.5361	0.3711	0.5258
Theil T: Eq Income	0.2371	0.7216	0.2474	0.7216	0.3402	0.5979	0.3093	0.6186
AROP: Percy	—	—	—	—	—	—	—	—
AROP: Eq Income	0.4845	0.4330	0.4639	0.4742	0.5567	0.3505	0.6082	0.2887
QSR: Percy	—	—	—	—	—	—	—	—
QSR: Eq Income	0.1649	0.7732	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000

Table 7.8: Non-coverage Probability of the 95% Standard Confidence Interval

By studying table. 7.8 it is clear that the overall performance of the standard confidence interval was quite poor. It seems as if the different weighting methods resulted in better coverage in the upper limit than in the case of no weighting. This was however not a consistent trend. None of the non-coverage probabilities were close to the desired $\alpha = 0.05$ significance level. Under-coverage occurred in both the upper and lower limits with values consistently larger than 0.05.

Table 7.9 contains the non-coverage probabilities of the percentile interval:

	None		Design		RR_pp		RR_ph	
WELFARE INDICES	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Headcount: Percy	0.0412	0.0928	0.0722	0.0309	0.0412	0.0206	0.0309	0.0206
Headcount: Eq Income	0.0103	0.0619	0.0309	0.0515	0.0000	0.0103	0.0000	0.0103
Pov Gap: Percy	0.0206	0.1134	0.0619	0.0619	0.0206	0.0412	0.0412	0.0412
Pov Gap: Eq Income	0.0412	0.0619	0.0515	0.0515	0.0000	0.0206	0.0206	0.0103
Pov Gap2: Percy	0.0103	0.1134	0.0309	0.0825	0.0206	0.0722	0.0206	0.0619
Pov Gap2: Eq Income	0.0412	0.0412	0.0515	0.0619	0.0000	0.0103	0.0103	0.0103
Gini: Percy	0.0000	0.3505	0.0000	0.3402	0.0103	0.0928	0.0103	0.0722
Gini: Eq Income	0.0000	0.2371	0.0000	0.2165	0.0000	0.0309	0.0000	0.0309
MLD: Percy	0.0000	0.3608	0.0000	0.3711	0.0103	0.1134	0.0206	0.0825
MLD: Eq Income	0.0000	0.2474	0.0000	0.2268	0.0000	0.0206	0.0000	0.0412
Theil T: Percy	0.0000	0.4227	0.0000	0.4124	0.0103	0.2577	0.0103	0.2371
Theil T: Eq Income	0.0000	0.3711	0.0000	0.3711	0.0000	0.1753	0.0000	0.1546
AROP: Percy	—	—	—	—	—	—	—	—
AROP: Eq Income	0.0515	0.0722	0.0412	0.0722	0.0000	0.0103	0.0000	0.0103
QSR: Percy	—	—	—	—	—	—	—	—
QSR: Eq Income	0.0000	0.2165	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000

Table 7.9: Non-coverage Probability of the 95% Percentile Confidence Interval

The percentile confidence interval exhibits much better accuracy. Most of the non-coverage probabilities compare quite well to $\alpha = 0.05$. Note that the confidence intervals based on RR_pp and RR_ph weighting show better coverage than no weight or design weight, although in some cases they do show signs of over-coverage. The lower limits show signs of over-coverage while the upper limits show signs of under-coverage. Overall this confidence interval shows better stability.

Table 7.10 contains the non-coverage probability of the bootstrap- t confidence interval:

	None		Design		RR_pp		RR_ph	
WELFARE INDICES	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Headcount: Percy	0.1134	0.1340	0.1856	0.0722	0.0722	0.0103	0.0309	0.0206
Headcount: Eq Income	0.0515	0.1237	0.4845	0.4742	0.5670	0.3814	0.6392	0.2990
Pov Gap: Percy	0.1237	0.1340	0.2062	0.1134	0.0619	0.0412	0.0619	0.0309
Pov Gap: Eq Income	0.0515	0.0928	0.4845	0.4639	0.5361	0.4227	0.5464	0.4227
Poverty Gap2: Percy	0.1340	0.1237	0.2268	0.0928	0.0412	0.0412	0.0515	0.0412
Poverty Gap2: Eq Income	0.0309	0.1134	0.4845	0.4639	0.5052	0.4021	0.5361	0.3814
Gini: Percy	0.0412	0.3505	0.0412	0.3505	0.0309	0.0825	0.0309	0.1031
Gini: Eq Income	0.0000	0.2062	0.2165	0.7732	0.4742	0.5155	0.4639	0.5155
MLD: Percy	0.0412	0.3299	0.0515	0.3299	0.0309	0.0825	0.0412	0.0825
MLD: Eq Income	0.0000	0.2165	0.1856	0.7835	0.4227	0.5052	0.4330	0.5052
Theil T: Percy	0.1134	0.4021	0.1031	0.3711	0.0206	0.1959	0.0206	0.1856
Theil T: Eq Income	0.0000	0.2990	0.2577	0.7320	0.3402	0.6082	0.3299	0.6289
AROP: Percy	—	—	—	—	—	—	—	—
AROP: Eq Income	0.1134	0.1753	0.4845	0.4742	0.5670	0.3711	0.6392	0.2990
QSR: Percy	—	—	—	—	—	—	—	—
QSR: Eq Income	0.0000	0.2577	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000

Table 7.10: Non-coverage Probability of the 95% Bootstrap- t Confidence Interval

Although the results in table 7.10 show that the bootstrap- t confidence interval is not as accurate as the percentile confidence interval, this confidence interval also shows better accuracy than the standard confidence interval. It does seem as if it is not adequate for the Laeken indices. The bootstrap- t confidence interval also shows better stability than the standard confidence interval. Signs of over-coverage and under-coverage do appear.

The second measure of accuracy to consider is the length of the respective confidence intervals. Refer to equation (7.17). These results are portrayed in figures 7.39 through 7.53. In all of these figures the standard confidence interval has the shortest length. This follows from the fact that in figures 7.31 through 7.38 it consistently resulted in the most conservative confidence intervals in comparison to the percentile and bootstrap- t confidence intervals. Its lengths and standardized lengths thus coincide with the poor non-coverage probability results seen in table 7.8.



Figure 7.39: 95% Standard Confidence Interval Length: Poverty Measures (Percy)

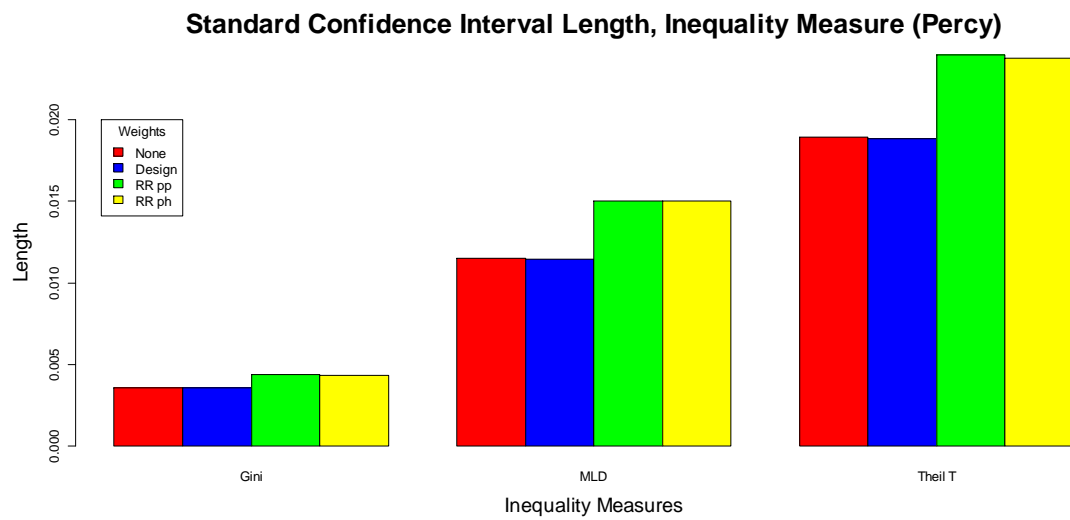


Figure 7.40: 95% Standard Confidence Interval Length: Inequality Measures (Percy)

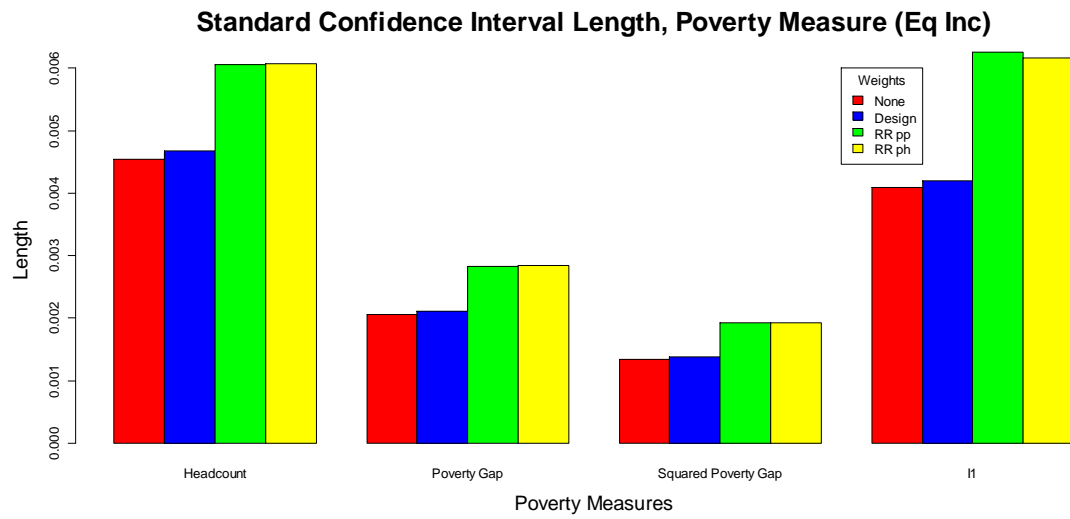


Figure 7.41: 95% Standard Confidence Interval Length: Poverty Measures (Eq Income)

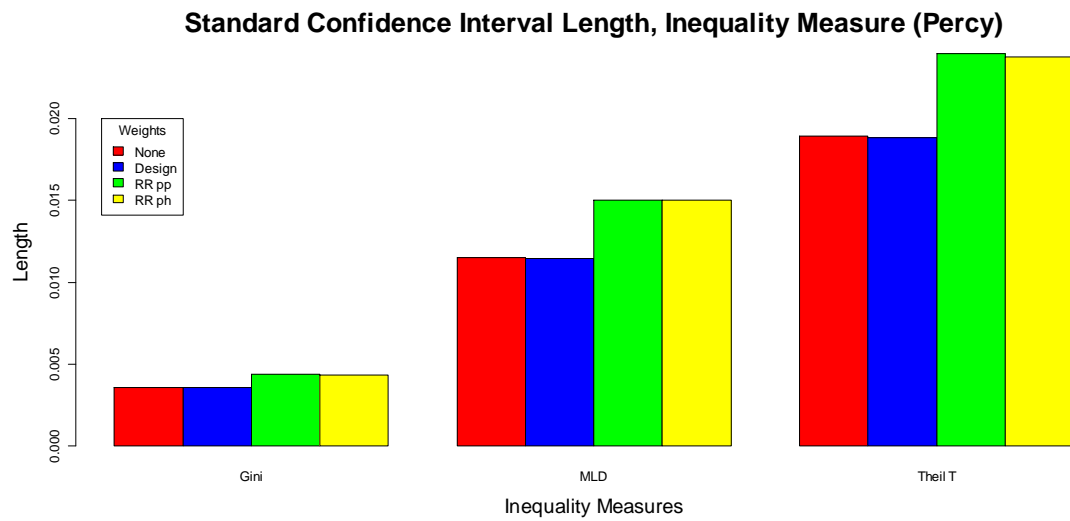


Figure 7.42: 95% Standard Confidence Interval Length: Inequality Measures (Eq Income)

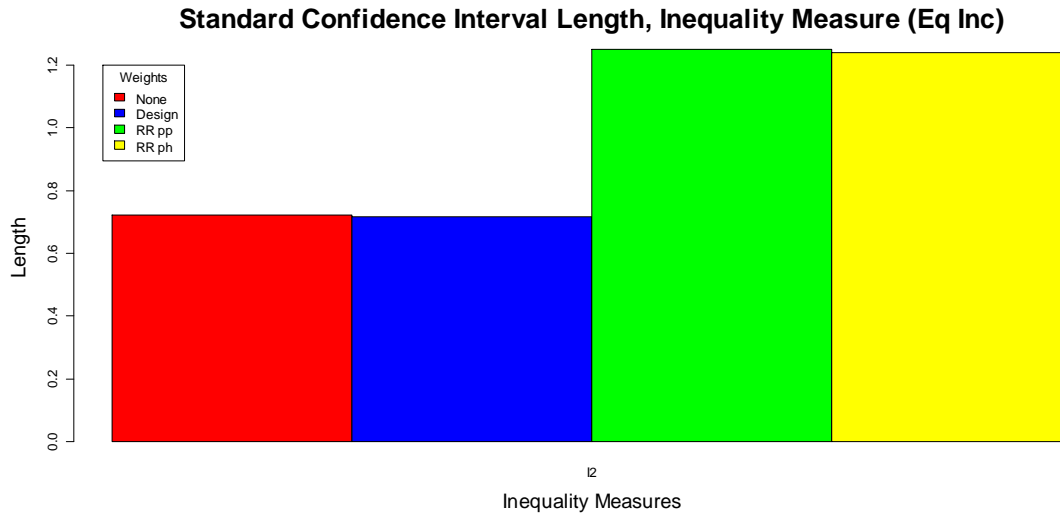


Figure 7.43: 95% Standard Confidence Interval Length: \hat{I}_2 (Eq Income)

By studying figures 7.39 through 7.43 the following conclusions can be made:

- The different weighting methods resulted in slightly longer confidence intervals. This is expected since weighting results in larger variances which would influence the lengths of the confidence intervals. Design weight compared quite well to no weight. The short lengths of the standard (asymptotic) interval coincide with its poor coverage probability.

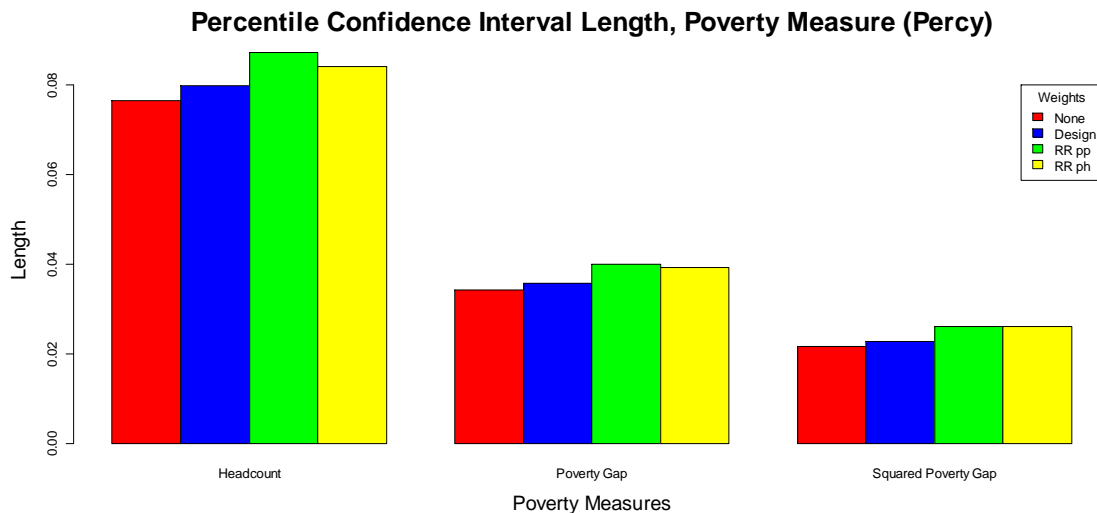


Figure 7.44: 95% Percentile Confidence Interval Length: Poverty Measures (Percy)

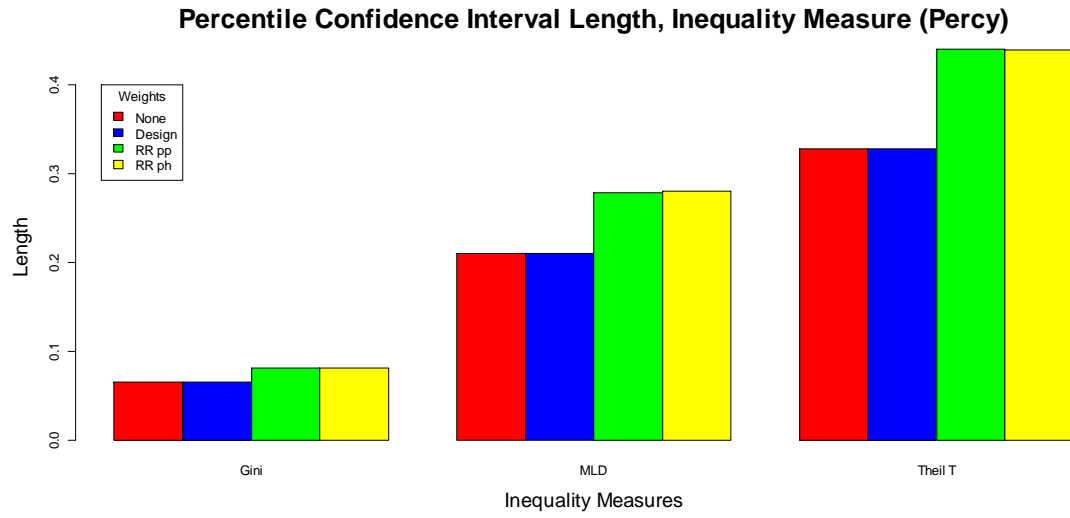


Figure 7.45: 95% Percentile Confidence Interval Length: Inequality Measures (Percy)

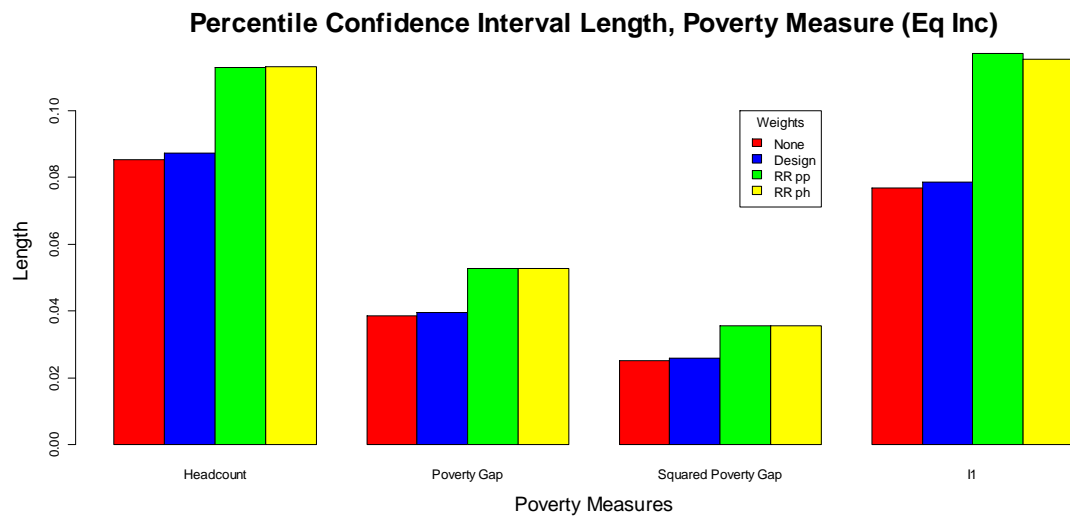


Figure 7.46: 95% Percentile Confidence Interval Length: Poverty Measures (Eq Income)

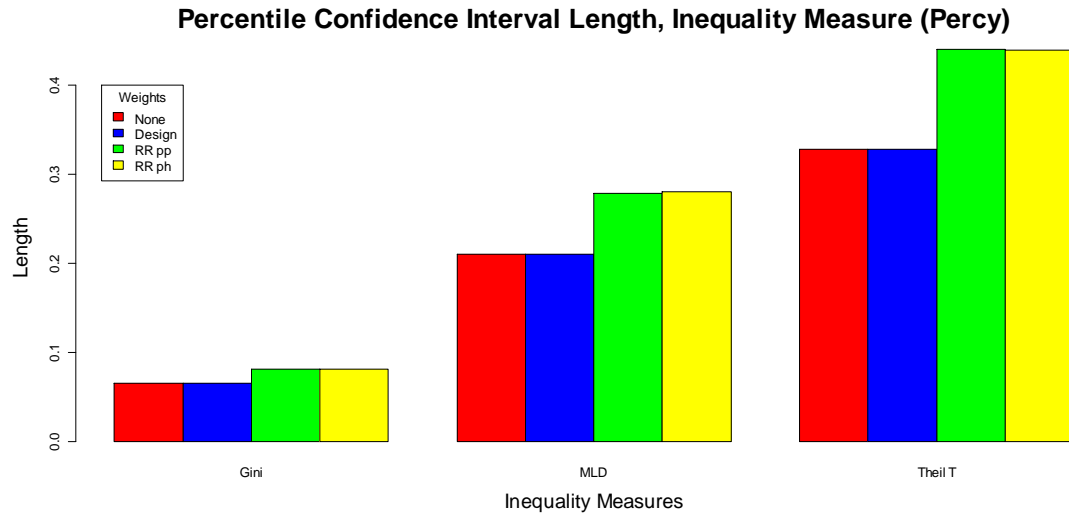


Figure 7.47: 95% Percentile Confidence Interval Length: Inequality Measures (Eq Income)

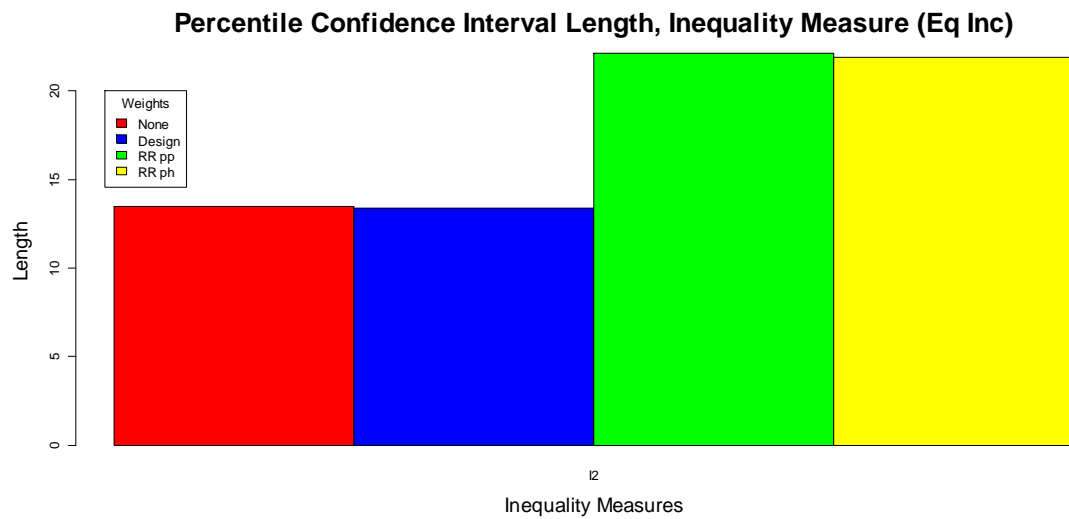


Figure 7.48: 95% Percentile Confidence Interval Length: \hat{I}_2 (Eq Income)

The following conclusions can be made about figures 7.44 through 7.48:

- As before, there are slight increases in the interval lengths over the different weighting methods.
- It appears as if the design weight and no weighting result in confidence intervals

with similar lengths while confidence intervals based on RR_pp and RR_ph have similar lengths.

- The increases are not so severe that weighting should be ignored as can be seen for many of the welfare indices.
- Note the short lengths of the percentile interval and yet its coverage probabilities were very good.

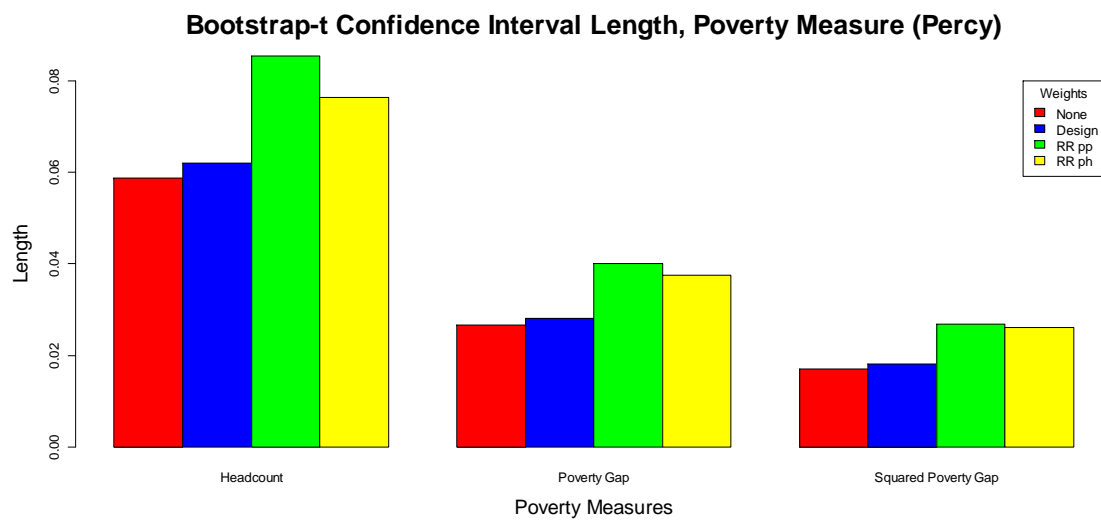


Figure 7.49: 95% Bootstrap- t Confidence Interval Length: Poverty Measures (Percy)

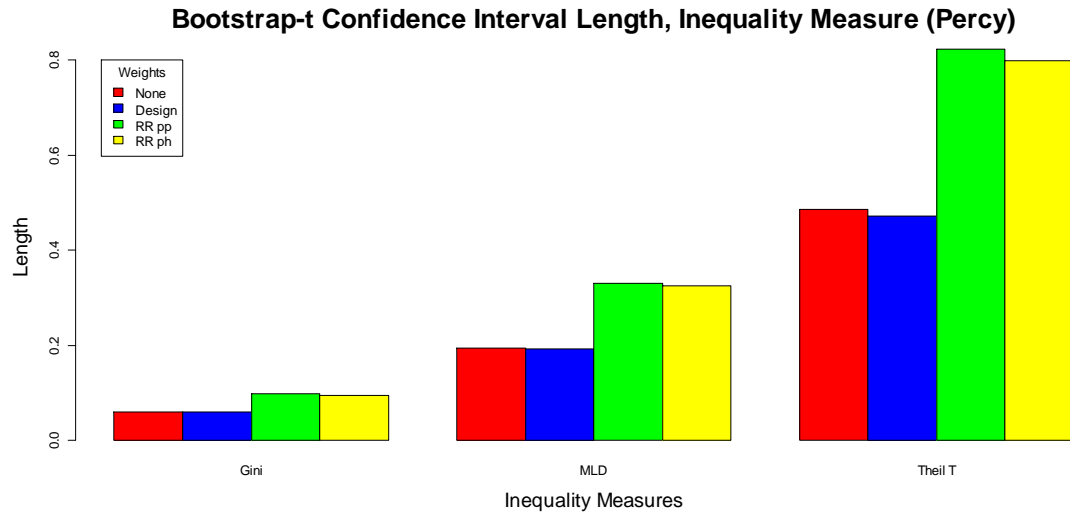


Figure 7.50: 95% Bootstrap- t Confidence Interval Length: Inequality Measures (Percy)

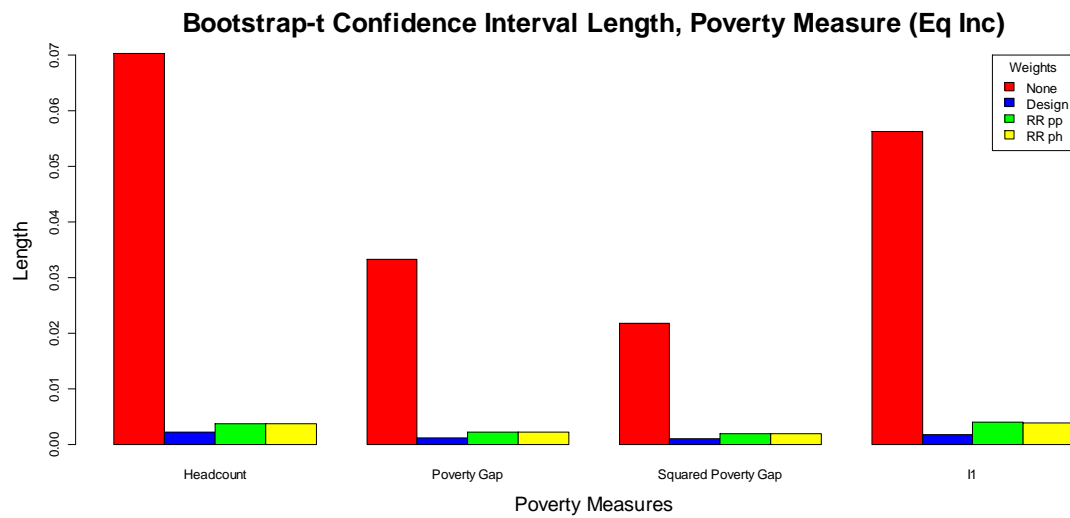


Figure 7.51: 95% Bootstrap- t Confidence Interval Length: Poverty Measures (Eq Income)

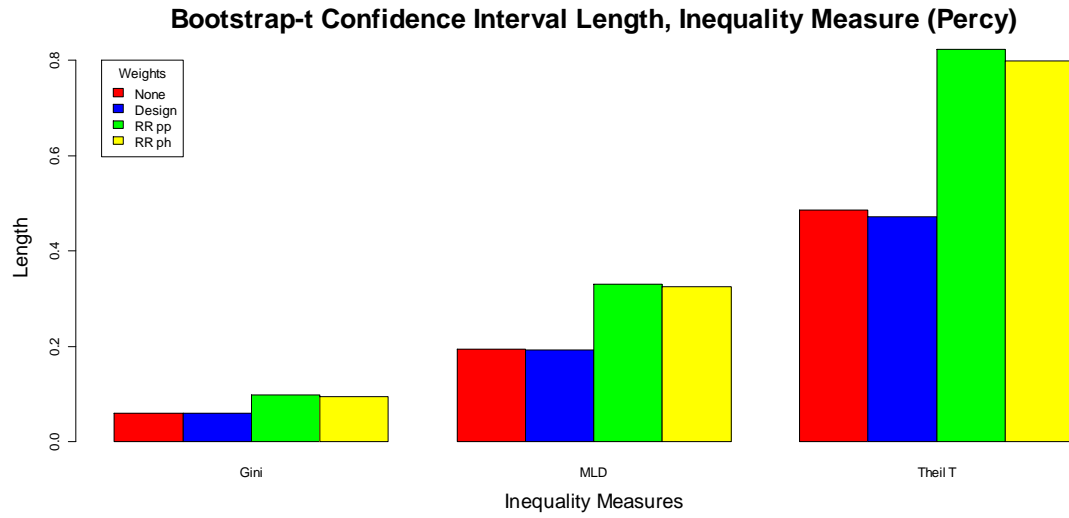


Figure 7.52: 95% Bootstrap- t Confidence Interval Length: Inequality Measures (Eq Income)

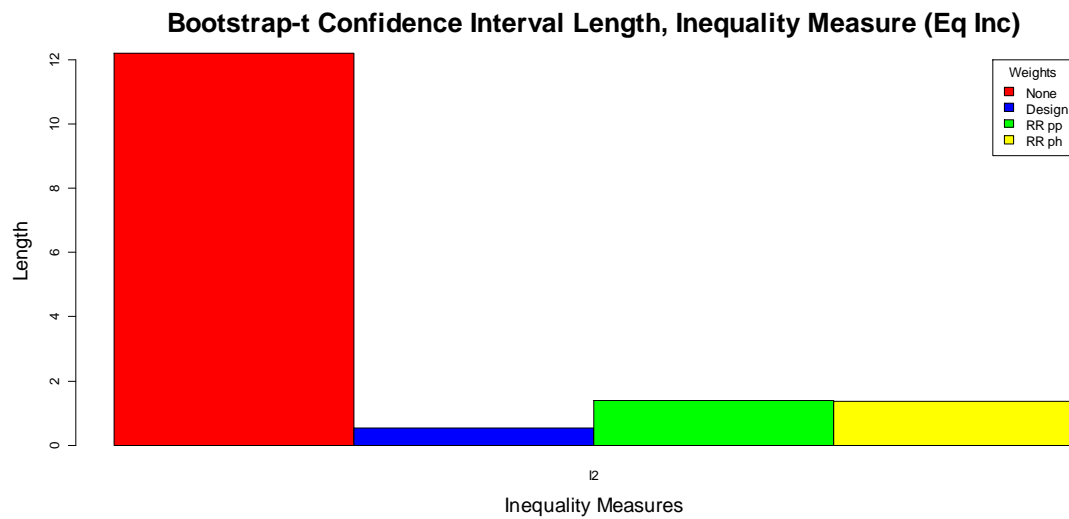


Figure 7.53: 95% Bootstrap- t Confidence Interval Length: \hat{I}_2 (Eq Income)

By studying figures 7.49 through 7.53 the following conclusions can be drawn:

- Results based on percy show similar trends to the previous two confidence intervals.
- In figures 7.51 and 7.53 it is seen that weighting improved upon the confidence

interval lengths in the case of equivalized income. The lengths seem shorter after weighting was applied.

- Figure 7.52 shows that design weight only slightly decreased the length obtained for no weighting, while RR_pp and RR_ph gave similar interval lengths. The lengths obtained under RR_ph were slightly shorter than under RR_pp, but longer than under no weighting and design weight.
- The short lengths of this interval coincide with it not having consistent good coverage probabilities.

In most of these figures weighting resulted in longer confidence intervals, except in the case of the bootstrap- t confidence interval which exhibited opposite results in some cases. Although not always desired, this certainly shows why the non-coverage probabilities in tables 7.8, 7.9 and 7.10 seem to improve as weighting was applied, especially for the absolute poverty line. Further study should however be carried out to determine why the results for the bootstrap- t confidence interval based on the relative poverty line differed from the other confidence intervals.

The third measure of accuracy to consider is the standardized lengths of the confidence intervals. The values of these graphs were calculated according to equation (7.18) and the results are shown in figures 7.54 through 7.59. Note that in these figures the standardized lengths of the percentile and bootstrap- t confidence intervals are larger than the standard confidence interval. It can be assumed that their improved coverage probabilities can be attributed to their longer standardized lengths.

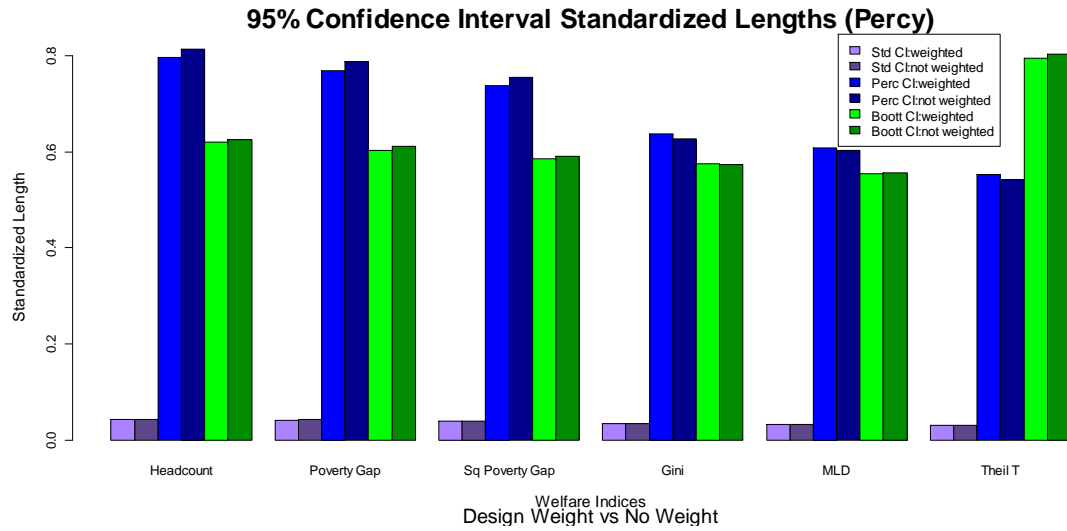


Figure 7.54: Comparison of Standardized Lengths of Confidence Intervals: Design Weight vs No Weight (Percy)

The following can be concluded from figure 7.54:

- It is clear that in the case of the design weight based on the absolute poverty line, weighting slightly improves the standardized length of the confidence interval for some welfare indices, but not all.
- The standard confidence interval has the smallest standardized length while the bootstrap- t confidence interval has a shorter standardized length than the percentile confidence interval, except for Theil T.

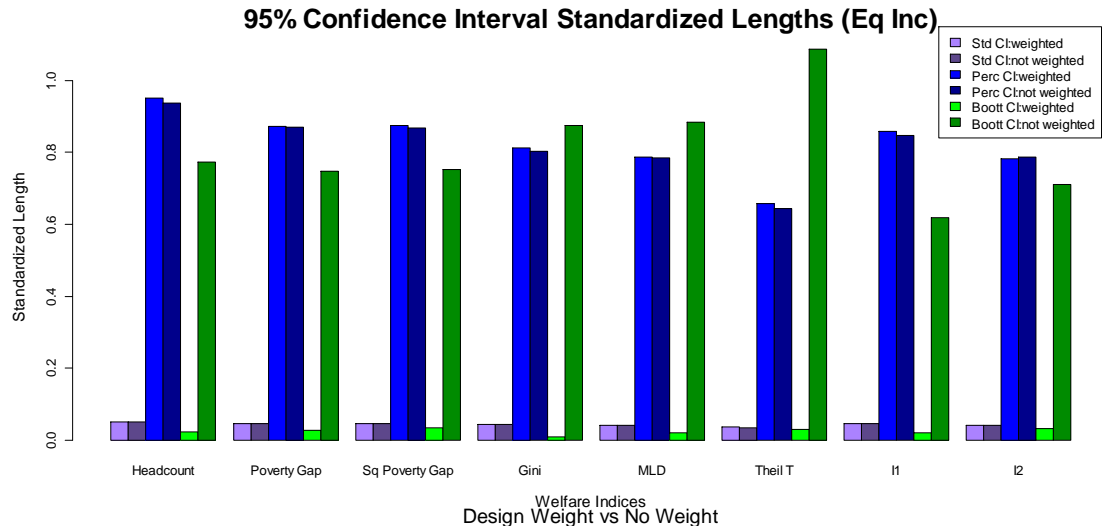


Figure 7.55: Comparison of Standardized Lengths of Confidence Intervals: Design Weight vs No Weight (Eq Income)

Conclusions that can be drawn from figure 7.55:

- In the case of the design weight based on the relative poverty line, weighting improved the standardized length of the bootstrap- t confidence interval.
- The other confidence intervals had slightly shorter standardized lengths under no weighting.

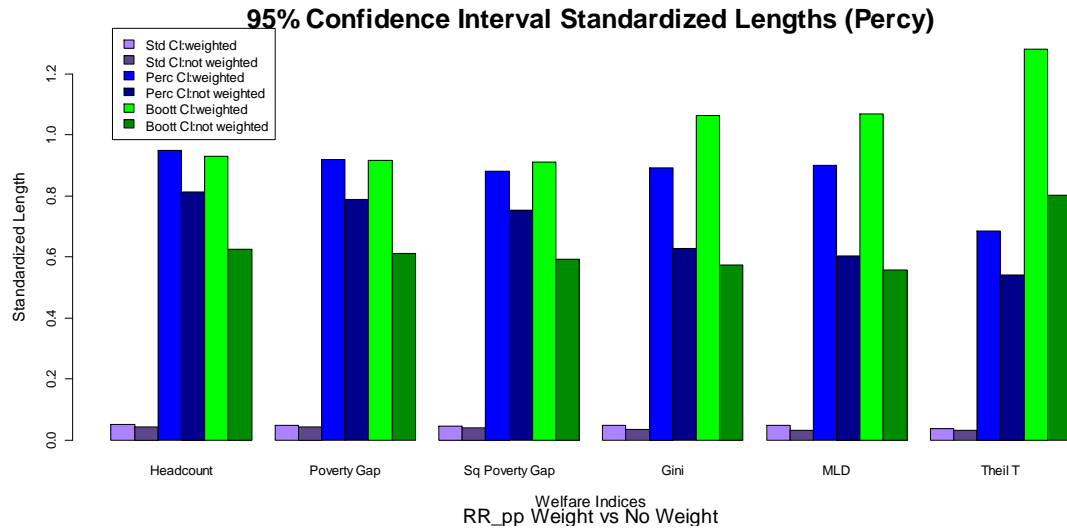


Figure 7.56: Comparison of Standardized Lengths of Confidence Intervals: RR_pp Weight vs No Weight (Percy)

By studying figure 7.56 the following conclusions can be drawn:

- Here it is seen that the RR_pp weighting, based on the absolute poverty line, resulted in longer standardized lengths than no weight.
- The unweighted standard confidence interval has the shortest standardized length followed by the unweighted bootstrap- t confidence interval.

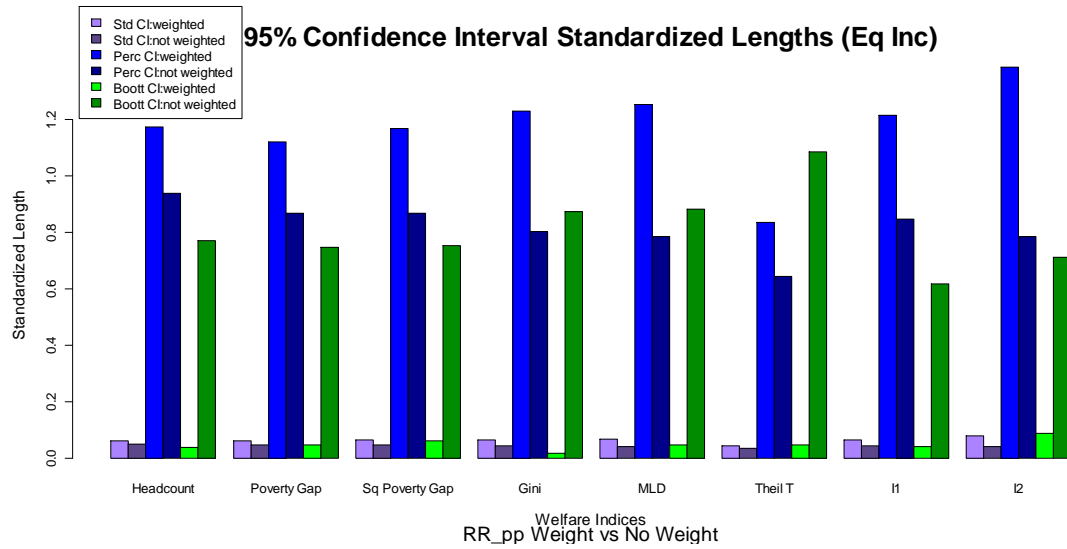


Figure 7.57: Comparison of Standardized Lengths of Confidence Intervals: RR_pp Weight vs No Weight (Eq Income)

The following conclusions can be drawn from figure 7.57:

- In the case of the standard confidence interval and the percentile confidence interval, based on the relative poverty line, RR_pp weighting resulted in longer standardized lengths.
- The weighted bootstrap- t confidence interval not only has the shortest standardized length, but in this case weighting resulted in it having shorter standardized lengths than for no weighting.

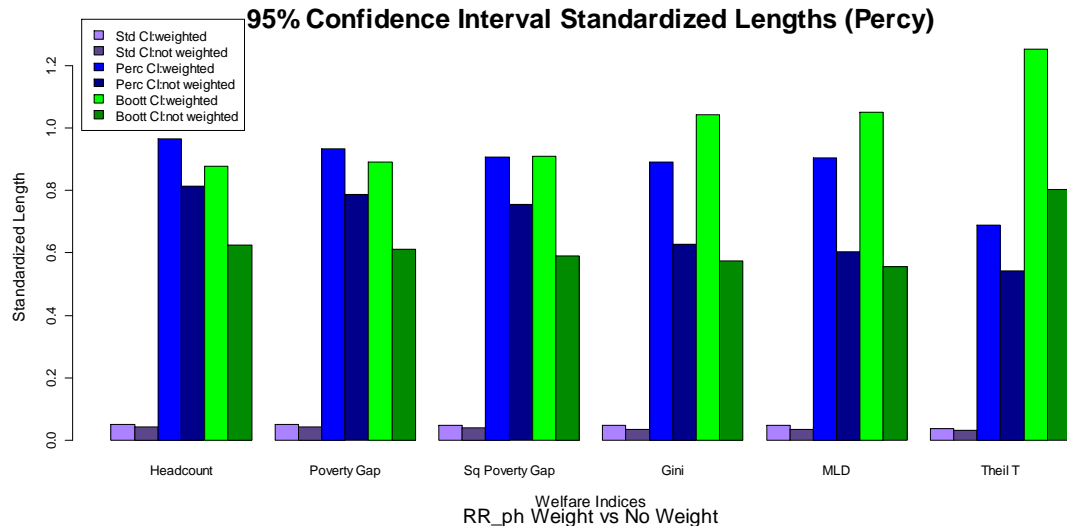


Figure 7.58: Comparison of Standardized Lengths of Confidence Intervals: RR_ph Weight vs No Weight (Percy)

By studying figure 7.58 the following conclusions can be drawn:

- Here similar results to those in figure 7.56 are seen.
- The unweighted standard confidence interval resulted in the shortest standardized length followed by the unweighted bootstrap- t confidence interval.

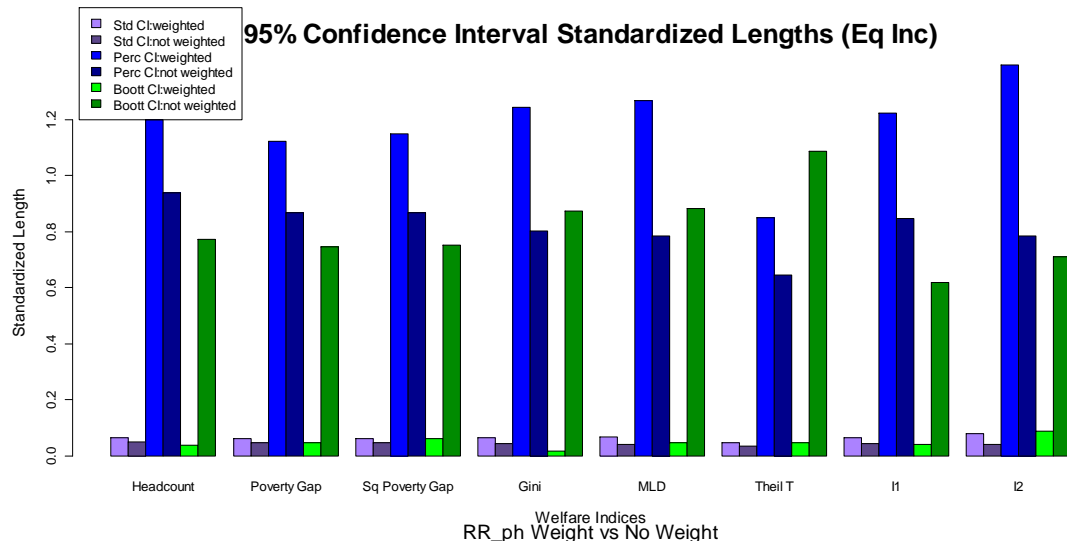


Figure 7.59: Comparison of Standardized Lengths of Confidence Intervals: RR_ph Weight vs No Weight (Eq Income)

The following can be concluded from figure 7.59:

- Similar results to figure 7.57.
- The bootstrap- t confidence interval based on the weighted values resulted in the shortest standardized length.

The results seen in the standardized length figures coincide with the non-coverage results and the length results obtained previously. Short standardized lengths went with poor coverage, and vice-versa.

The last measure of accuracy to consider is the shape of the confidence intervals. Refer to equation (7.19) to see how the values were calculated. As will be seen in figures 7.60 and 7.61, the shape of the standard confidence interval for all welfare measures calculated with all weighting methods and based on the absolute poverty line, is equal to 1. The same result was found when the relative poverty line was used, but it is not shown here. This is as expected since the standard confidence interval is always symmetric about the parameter of interest. Note that the black horizontal line in figures 7.60 through 7.69 is drawn at the shape value equal to one so as to compare the shape of the confidence intervals to the shape of the standard confidence interval.

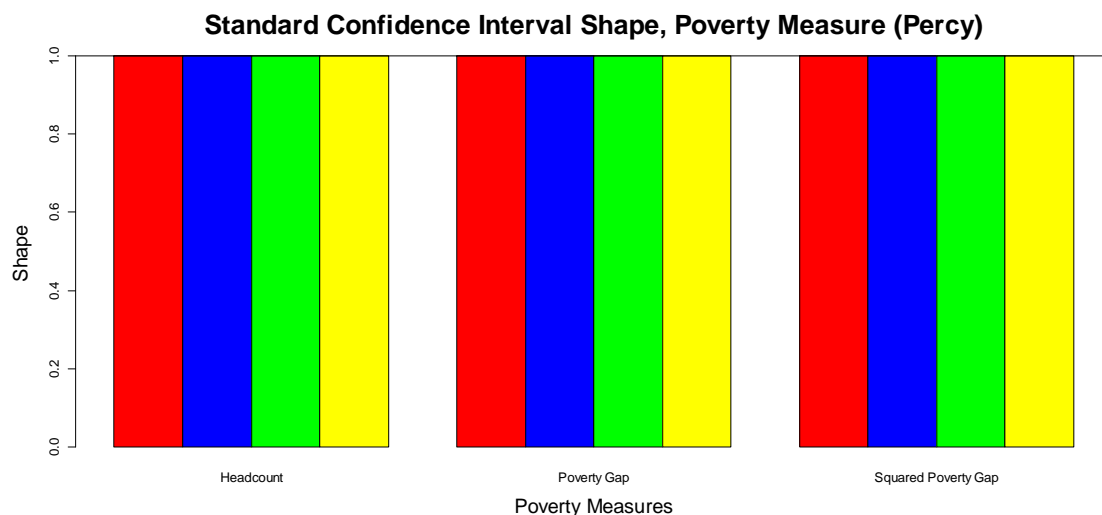


Figure 7.60: 95% Standard Confidence Interval Shape: Poverty Measures (Percy)

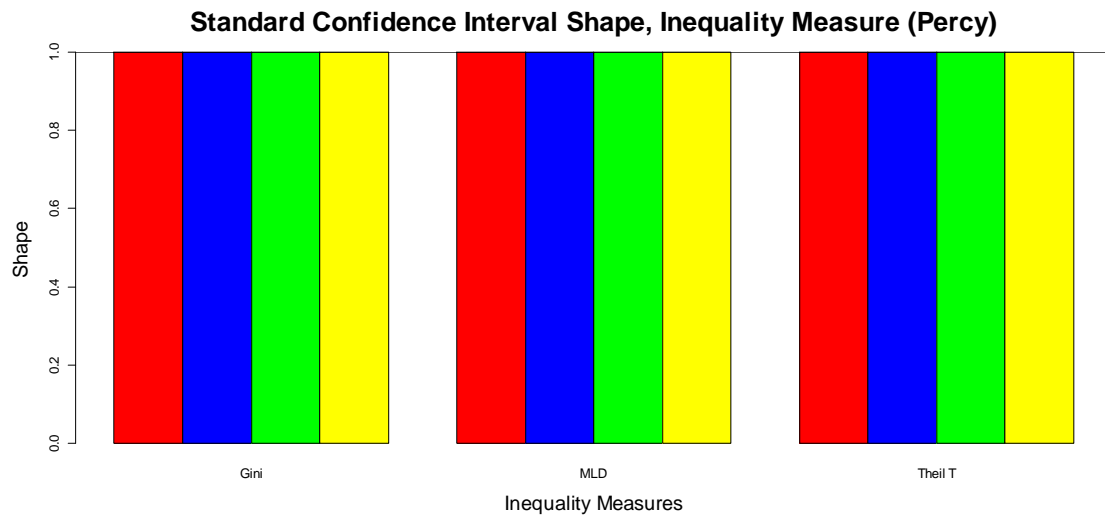


Figure 7.61: 95% Standard Confidence Interval Shape: Inequality Measures (Percy)

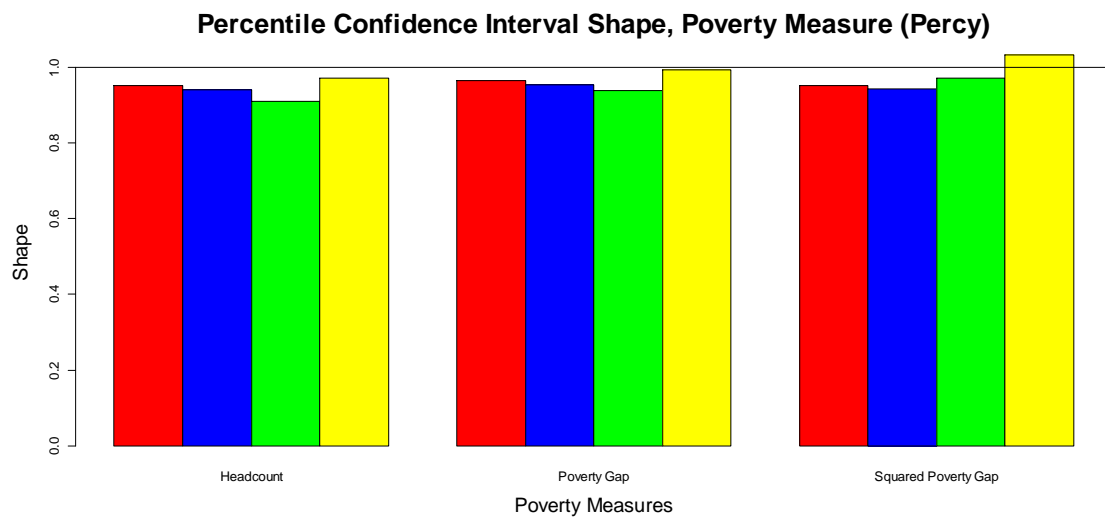


Figure 7.62: 95% Percentile Confidence Interval Shape: Poverty Measures (Percy)

In figure 7.62 it can be seen how close the percentile confidence interval shape is to 1. It generally is smaller than one, indicative of the lower limit of the interval over all weighting methods being further away from the value of the parameter of interest than the upper limit. Thus, it tends to be skew to the left. The weighting methods, however, slightly increase the shape of the percentile interval to 1.

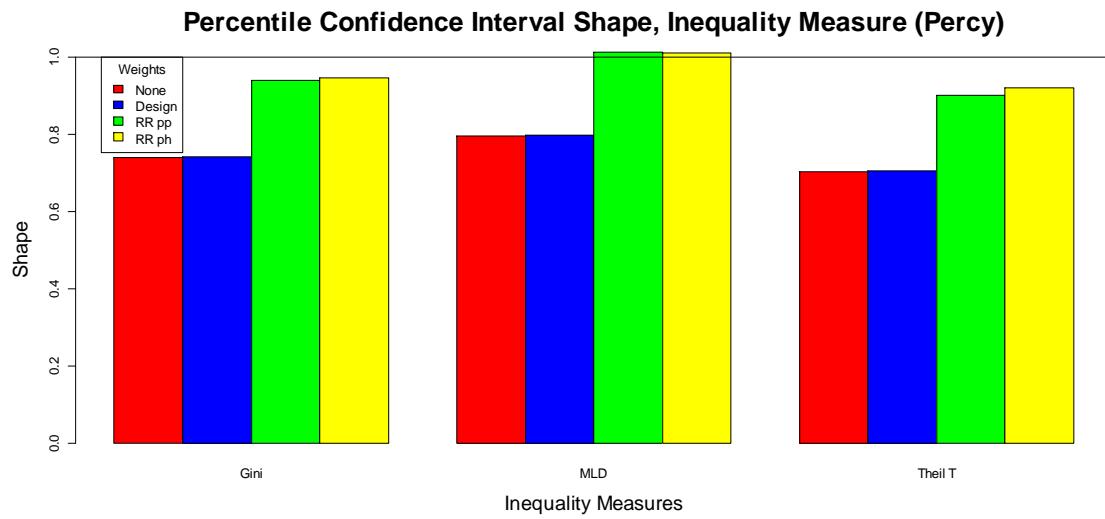


Figure 7.63: 95% Percentile Confidence Interval Shape: Inequality Measures (Percy)

Figure 7.63, containing the shapes of the percentile confidence intervals for the inequality measures based on the absolute poverty line, shows an even further increase in the shape of the intervals to 1 as weighting is applied. Once again it is apparent that the percentile confidence interval tends to be skew to the left.

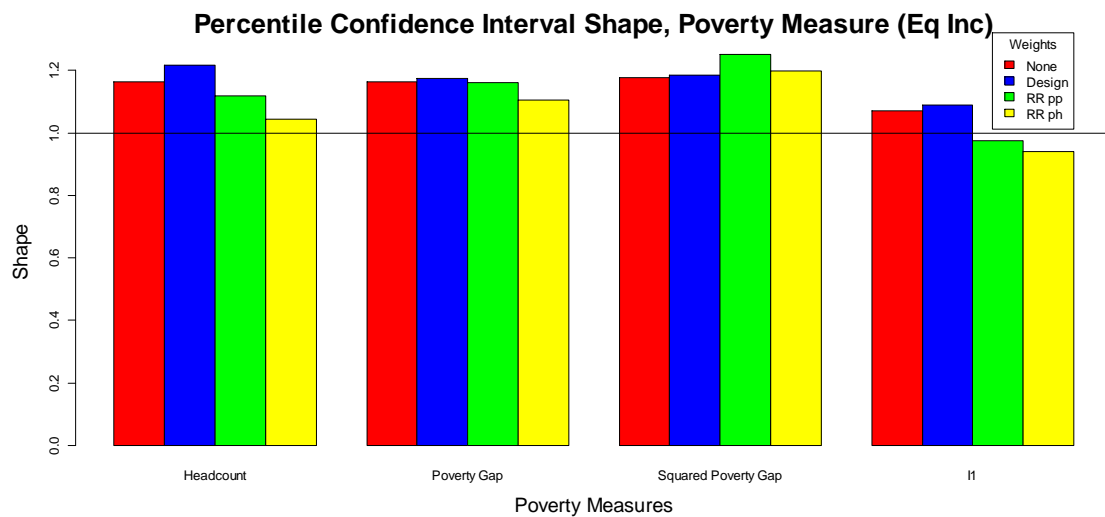


Figure 7.64: 95% Percentile Confidence Interval Shape: Poverty Measures (Eq Income)

In figure 7.64 it is clear the the shape of the percentile confidence intervals for the poverty measures based on equivalised income tends to be larger than one. The shape decreases to 1 as weighting is applied showing the “similarity” of the percentile confidence intervals and the standard confidence intervals.

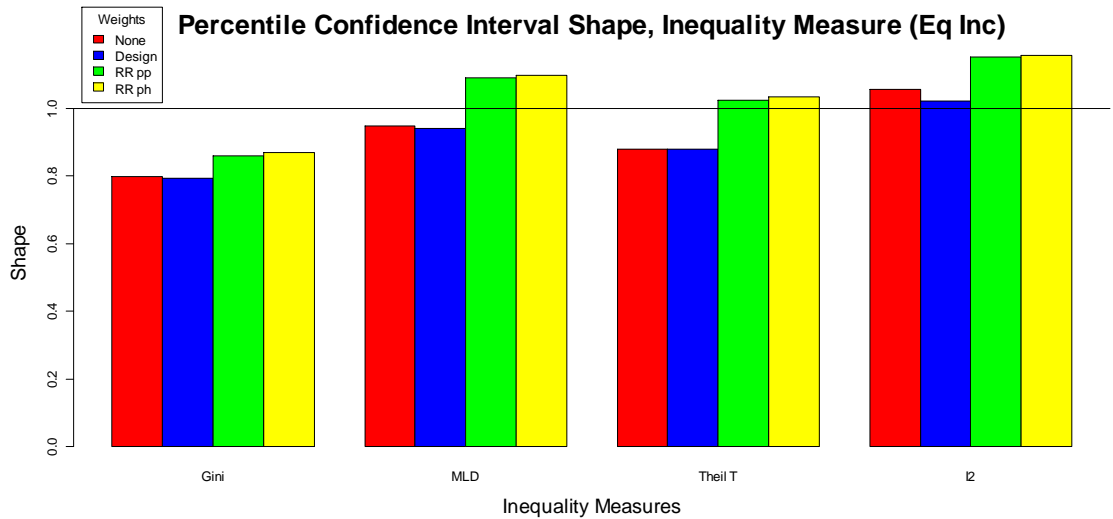


Figure 7.65: 95% Percentile Confidence Interval Shape: Inequality Measures (Eq Income)

From figure 7.65 similar conclusions about the shape of the percentile confidence interval can be drawn as from the previous figures. The only difference here is that the confidence interval seems to be more skew to the left with some of the estimated shapes lying further below the horizontal line.

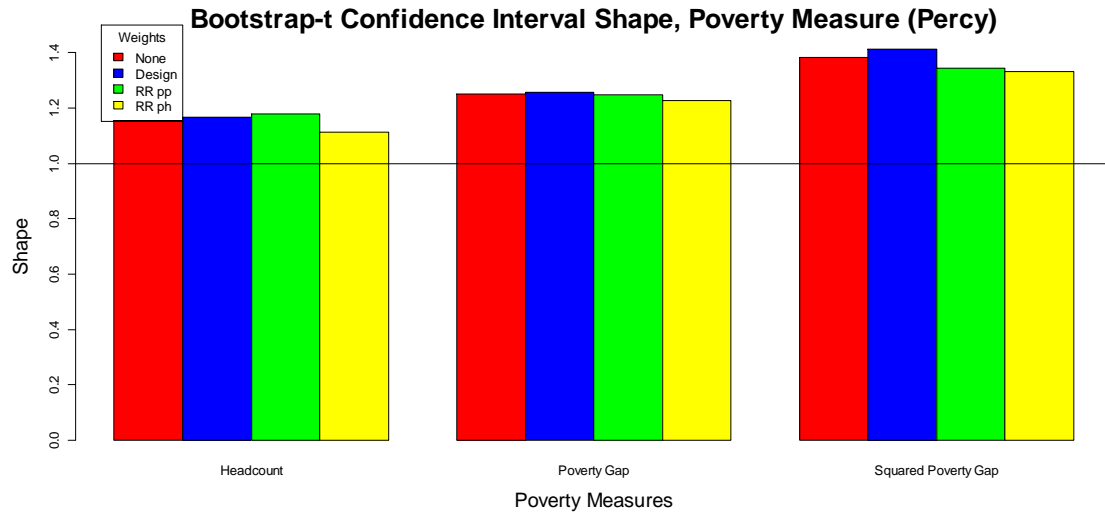


Figure 7.66: 95% Bootstrap- t Confidence Interval Shape: Poverty Measures (Percy)

From the theory about the bootstrap- t confidence interval, it is known that it tends to take the skewness of the data it is based on, into account. In figure 7.66 this is seen. As noted before in figures 7.31 through 7.38, there seems to be a skewness to the right in the distributions of the welfare indices. This can only be because of similar skewness in the data. In this figure it is seen that the shape of the bootstrap- t confidence interval is larger than one, indicative of the upper limit of the confidence interval being farther away from the value of the parameter of interest, than the lower limit. Weighting once again slightly improved upon the shape of the confidence interval, with the shape of the confidence interval based on RR_ph weighting being the closest to 1.

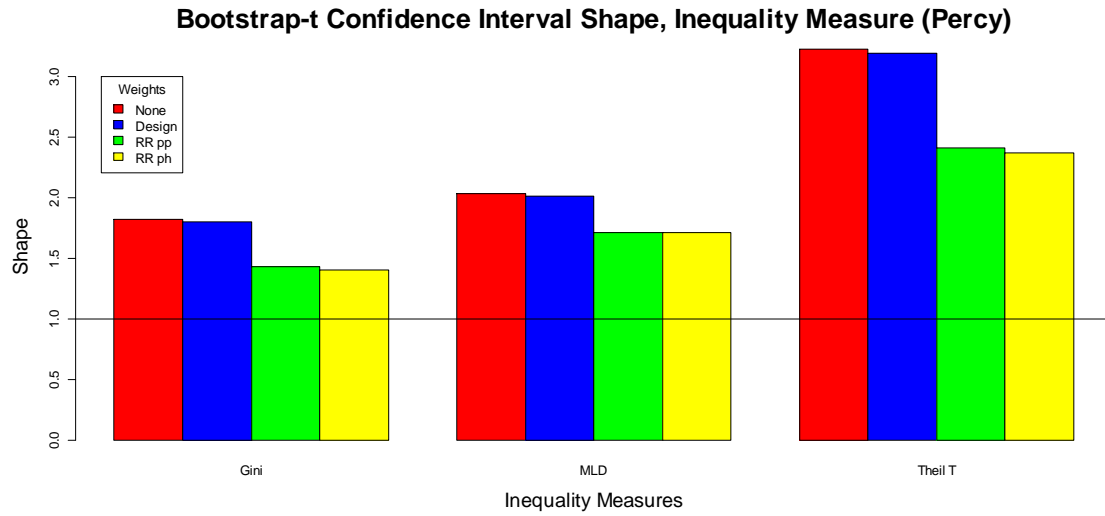


Figure 7.67: 95% Bootstrap-*t* Confidence Interval Shape: Inequality Measures (Percy)

Figure 7.67, containing results for the inequality measures based on the absolute poverty line, shows a similar pattern than figure 7.66. The bootstrap-*t* confidence interval based on RR_ph weighting having a shape closest to 1.

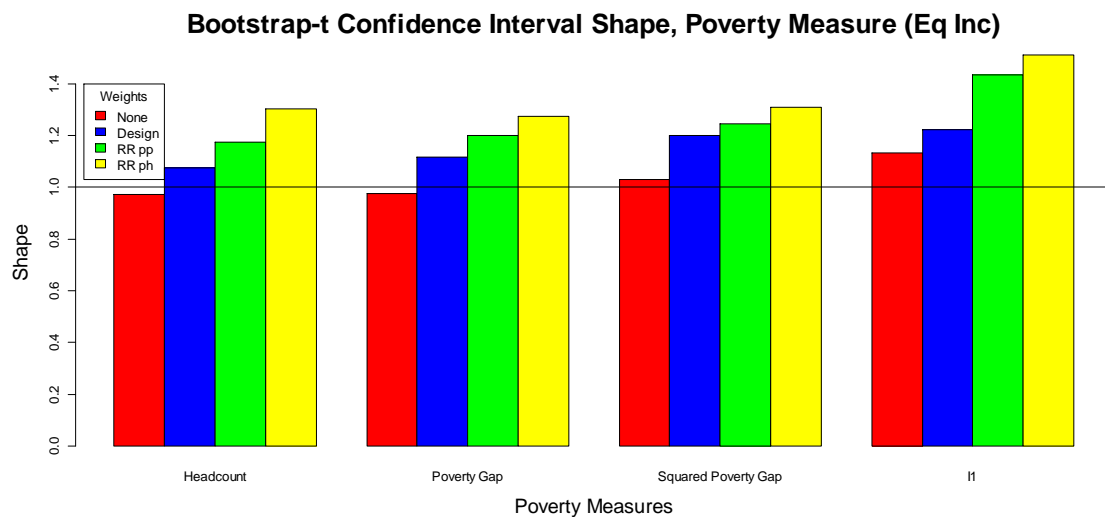


Figure 7.68: 95% Bootstrap-*t* Confidence Interval Shape: Poverty Measures (Eq Income)

Figure 7.68 with results for the poverty measures based on the relative poverty line, is the only figure for the bootstrap-*t* confidence interval that does not improve

upon the shape as weighting is applied. For this figure it must be concluded that no weighting had resulted in a confidence interval with shape closest to 1.

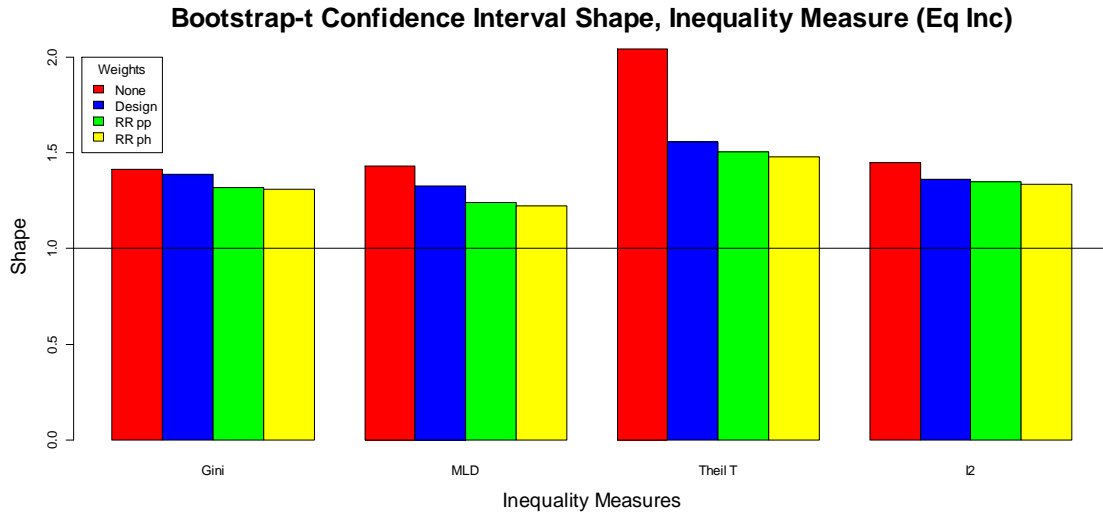


Figure 7.69: 95% Bootstrap- t Confidence Interval Shape: Inequality Measures (Eq Income)

Figure 7.69 once again shows results similar to figures 7.66 and 7.67, namely that the bootstrap- t confidence interval is skew to the right and that weighting improved the shape of the confidence interval. Weighting method RR_ph resulted in a confidence interval with shape closest to 1.

Note that in this simulation study, where non-parametric resampling methods were used, there does in essence not exist a “golden standard” for comparison of the confidence intervals. Thus to compare confidence intervals, it is best to make conclusions based on their non-coverage probabilities, as well as their lengths. These two measures are closely related, since a shorter interval length typically implies larger non-coverage probability and vice-versa.

7.3 Summary and Conclusions

The purpose of the simulation study can be summarised as twofold:

1. to investigate the advantages of weighting for inferences in the case of complex sampling, and

2. to show how reliable the use of resampling methods is for inference in complex sampling.

In the previous section the results of the simulation study were given in graphical form. For more convenient drawing of conclusions, in terms of the above purpose, the results are now given in tabular form.

This section consists of two parts. First, poverty measures are considered in tables with regard to the absolute poverty line and the relative poverty line, after which the results for the inequality measures are given. In each table the “best” value is given in green. The values given in brackets represent the standard error for that estimate under bootstrap resampling.

Poverty Measures									
Absolute Poverty Line									
Headcount									
PERFORMANCE FACTORS		None		Design		RR_pp		RR_ph	
True Bias		-0.0013		0.0069		0.0032		-0.0003	
True MSE		0.0006		0.0007		0.0005		0.0005	
Dev_{bias}		0.0014 (0.00010)		-0.0068 (0.00010)		-0.0052 (0.00014)		0.0000 (0.00020)	
Dev_{MSE}		0.0000 (0.00001)		-0.0001 (0.00001)		0.0002 (0.00001)		0.0002 (0.00001)	
% Rel. Bias: $\widehat{var}(\hat{\theta})$		-4.3104		-1.3474		31.5594		34.3294	
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$		-4.5702		-8.5286		29.8195		34.3192	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.433	0.4742	0.5773	0.3505	0.5258	0.3918	0.4845	0.4536
	Perc	0.0412	0.0928	0.0722	0.0309	0.0412	0.0206	0.0309	0.0206
	Boott	0.1134	0.134	0.1856	0.0722	0.0722	0.0103	0.0309	0.0206
Length:	Std	0.0041		0.0043		0.0046		0.0045	
	Perc	0.0764		0.0798		0.0872		0.0841	
	Boott	0.0588		0.062		0.0853		0.0764	
Std Length:	Std	0.0435		0.0426		0.0506		0.0514	
	Perc	0.8137		0.7972		0.9495		0.9669	
	Boott	0.6257		0.62		0.9294		0.878	
Shape:	Std	1		1		1		1	
	Perc	0.9512		0.9399		0.9093		0.9711	
	Boott	1.1563		1.1663		1.1803		1.1123	

Table 7.11: Summary of performance measures calculated for headcount based on the absolute poverty line

The performance factors calculated for the headcount estimator seemed to be improved by the application of RR_ph. Small percentage relative biases were found for the application of no weight and design weight with the values for RR_pp and RR_ph being quite large. Non-coverage probability also seemed to be slightly improved by weighting with RR_ph, giving non-coverage values closest to 0.025 in each tail. The percentile confidence interval seemed to be the best suited confidence interval for the headcount index, while the standard (asymptotic) confidence interval did not give good results. Its short length is a result of its very poor coverage probability, as can be seen from table (7.11). The percentile interval length was slightly larger in comparison to the standard and bootstrap- t confidence intervals, hence the better coverage probabilities. Performance factors calculated for the confidence intervals showed slight improvement under weighting with design weight and RR_ph. The standard errors calculated for Dev_{bias} and Dev_{MSE} were all very small. The bootstrap resampling could be regarded as reliable for the estimation of the headcount index in this case.

Poverty Measures									
Absolute Poverty Line									
Poverty Gap									
PERFORMANCE FACTORS		None	Design	RR_pp	RR_ph				
True Bias		-0.0002	0.0032	0.0020	0.0008				
True MSE		0.0001	0.0001	0.0001	0.0001				
Dev_{bias}		0.0003 (0.00005)	-0.0031 (0.00005)	-0.0027 (0.00006)	-0.0008 (0.00010)				
Dev_{MSE}		0.0000 (0.00000)	0.0000 (0.00000)	0.0000 (0.00000)	0.0000 (0.00000)				
% Rel. Bias: $\widehat{var}(\hat{\theta})$		-9.9025	-7.1789	25.2175	27.6455				
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$		-9.9145	-13.9054	21.5690	26.9873				
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.4536	0.4948	0.5361	0.3814	0.5155	0.4124	0.4948	0.4227
	Perc	0.0206	0.1134	0.0619	0.0619	0.0206	0.0412	0.0412	0.0412
	Boott	0.1237	0.134	0.2062	0.1134	0.0619	0.0412	0.0619	0.0309
Length:	Std	0.0018		0.0019		0.0021		0.0021	
	Perc	0.0342		0.0358		0.0401		0.0393	
	Boott	0.0266		0.028		0.04		0.0375	
Std Length:	Std	0.0421		0.0412		0.0489		0.0499	
	Perc	0.7879		0.7699		0.9184		0.9325	
	Boott	0.612		0.6023		0.9176		0.8914	
Shape:	Std	1		1		1		1	
	Perc	0.9641		0.9535		0.9377		0.9927	
	Boott	1.252		1.257		1.2486		1.226	

Table 7.12: Summary of performance measures calculated for poverty gap based on the absolute poverty line

Table 7.12 shows that no weighting worked best for the poverty gap index. Also design weight gave pleasing results. Standard errors under bootstrap resampling are still very small, thus indicative of the reliability of the bootstrap method. Percentage relative bias under design weight and no weight are quite small once again. Non-coverage of the standard (asymptotic) interval was quite poor, with the percentile interval performing much better. Although its total non-coverage was not exactly 0.05, the weighting seemed to improve upon it, with RR_pp weighting resulting in total non-coverage of approximately 0.06. The poor coverage of the standard interval coincides with it having the smallest length once again, but the percentile and bootstrap- t confidence intervals' lengths are not much larger. Weighting resulted in slightly longer confidence interval lengths, but design weight gave slight improvement

to the standardized lengths.

Poverty Measures									
Absolute Poverty Line									
Squared Poverty Gap									
PERFORMANCE FACTORS		None		Design		RR_pp		RR_ph	
True Bias		0.0000		0.0018		0.0012		0.0006	
True MSE		0.0001		0.0001		0.0001		0.0001	
Dev_{bias}		0.0001 (0.00003)		-0.0017 (0.00003)		-0.0016 (0.00004)		-0.0006 (0.00006)	
Dev_{MSE}		0.0000 (0.00000)		0.0000 (0.00000)		0.0000 (0.00000)		0.0000 (0.00000)	
% Rel. Bias: $\widehat{var}(\hat{\theta})$		-16.7368		-15.9829		15.7698		20.7832	
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$		-16.7252		-20.4204		12.9208		20.0397	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.4433	0.5052	0.5052	0.4227	0.4845	0.4227	0.4536	0.4845
	Perc	0.0103	0.1134	0.0309	0.0825	0.0206	0.0722	0.0206	0.0619
	Boott	0.134	0.1237	0.2268	0.0928	0.0412	0.0412	0.0515	0.0412
Length:	Std	0.0012		0.0012		0.0014		0.0014	
	Perc	0.0217		0.0228		0.0261		0.0261	
	Boott	0.017		0.0181		0.0269		0.0261	
Std Length:	Std	0.0404		0.0394		0.047		0.0484	
	Perc	0.7544		0.738		0.8816		0.9071	
	Boott	0.5917		0.586		0.9096		0.9099	
Shape:	Std	1		1		1		1	
	Perc	0.952		0.9427		0.9716		1.0319	
	Boott	1.3822		1.4127		1.3454		1.3331	

Table 7.13: Summary of performance measures calculated for squared poverty gap based on the absolute poverty line

The estimation of the squared poverty gap achieved best results under no weighting and design weight. Once again the reliability of the bootstrap method is seen through the very small standard errors of Dev_{bias} and Dev_{MSE} . Slightly larger percentage relative bias values than in the previous tables were obtained, but here the values seemed to be quite similar over all the weighting methods, but best under RR_pp weighting. The percentile confidence interval gave best coverage and with weighting applied, these values moved closer to 0.05. Its best non-coverage, approximately 0.08, was achieved with RR_ph. The poor coverage of the standard (asymptotic) interval coincides with its short length. Percentile and bootstrap- t

confidence intervals have similar lengths, not very long, with their best length being achieved under no weighting. Design weight resulted in the best standardized lengths.

Poverty Measures									
Relative Poverty Line									
Headcount									
PERFORMANCE FACTORS		None		Design		RR_pp		RR_ph	
True Bias		-0.0021		-0.0028		0.0013		0.0038	
True MSE		0.0005		0.0005		0.0006		0.0006	
Dev_{bias}		0.0028 (0.00063)		0.0046 (0.00059)		-0.0005 (0.00064)		-0.0049 (0.00075)	
Dev_{MSE}		0.0002 (0.00002)		0.0002 (0.00002)		0.0006 (0.00002)		0.0007 (0.00002)	
% Rel. Bias: $\widehat{var}(\hat{\theta})$		34.6613		39.1125		105.5913		122.8276	
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$		33.6163		37.7684		105.1380		117.5128	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.4742	0.4227	0.4639	0.4742	0.5464	0.3505	0.6186	0.299
	Perc	0.0103	0.0619	0.0309	0.0515	0	0.0103	0	0.0103
	Boott	0.0515	0.1237	0.4845	0.4742	0.567	0.3814	0.6392	0.299
Length:	Std	0.0045		0.0047		0.0061		0.0061	
	Perc	0.0852		0.0873		0.113		0.1133	
	Boott	0.0702		0.0022		0.0036		0.0037	
Std Length:	Std	0.05		0.0509		0.0628		0.0643	
	Perc	0.9375		0.9511		1.1731		1.1989	
	Boott	0.7721		0.0239		0.0377		0.0388	
Shape:	Std	1		1		1		1	
	Perc	1.1642		1.2176		1.1197		1.0449	
	Boott	0.973		1.0759		1.1737		1.3028	

Table 7.14: Summary of performance measures calculated for headcount based on the relative poverty line

From table 7.14 it is clear that no weighting achieved best results for estimating the headcount index. The standard errors for estimating Dev_{bias} and Dev_{MSE} were very small, indicative of the reliability of the bootstrap method. Somewhat larger percentage relative bias values than previously achieved were obtained. No weighting and design weight gave similar values, but RR_pp and RR_ph achieved quite large percentage relative biases. The best non-coverage was achieved by the percentile confidence interval with some improvement under the weighting meth-

ods. Both RR_pp and RR_ph gave the smallest total non-coverage, approximately 0.01, which suggests over-estimation. Design weight achieved total non-coverage of approximately 0.08, also not significantly different from 0.05. The standard (asymptotic) confidence interval once again achieved shortest length and hence it gave the poorest non-coverage as well. The bootstrap- t confidence interval seemed to be shorter than the percentile confidence interval under the weighting methods. This could have contributed to it not achieving quite the same non-coverage as the percentile confidence interval.

Poverty Measures									
Relative Poverty Line									
Poverty Gap									
PERFORMANCE FACTORS		None		Design		RR_pp		RR_ph	
True Bias		0.0004		0.0001		0.0020		0.0031	
True MSE		0.0001		0.0001		0.0001		0.0001	
Dev_{bias}		0.0002 (0.00016)		0.0005 (0.00017)		-0.0017 (0.00019)		-0.0033 (0.00023)	
Dev_{MSE}		0.0000 (0.00000)		0.0000 (0.00000)		0.0001 (0.00001)		0.0001 (0.00001)	
% Rel. Bias: $\widehat{var}(\hat{\theta})$		10.6175		12.3000		91.8397		101.7201	
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$		10.7526		12.5729		86.3054		88.4295	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.4742	0.4536	0.4536	0.4536	0.5155	0.4021	0.5155	0.4227
	Perc	0.0412	0.0619	0.0515	0.0515	0	0.0206	0.0206	0.0103
	Boott	0.0515	0.0928	0.4845	0.4639	0.5361	0.4227	0.5464	0.4227
Length:	Std	0.0021		0.0021		0.0028		0.0028	
	Perc	0.0386		0.0396		0.0527		0.0528	
	Boott	0.0332		0.0012		0.0022		0.0022	
Std Length:	Std	0.0463		0.0467		0.0603		0.0605	
	Perc	0.8691		0.872		1.1222		1.1216	
	Boott	0.747		0.0271		0.0461		0.0459	
Shape:	Std	1		1		1		1	
	Perc	1.1636		1.1742		1.1605		1.1063	
	Boott	0.976		1.117		1.2003		1.2733	

Table 7.15: Summary of performance measures calculated for poverty gap based on the relative poverty line

Once again it is seen in table 7.15 that no weighting gave the best results for the estimation of the poverty gap. The standard errors calculated are very close

to zero indicating that the bootstrap method seems to be a reliable method for estimating this measure. Percentage relative biases under no weighting and design weight are similar and quite small, but RR_pp and RR_ph resulted in quite large percentage relative biases. The percentile confidence interval achieved a total non-coverage of approximately 0.03 under RR_ph, which is not significantly different from 0.05. The bootstrap- t confidence interval gave a satisfactory result under no weighting, but weighting worsened its non-coverage, which can also be seen from its short lengths achieved under the weighting methods.

Poverty Measures									
Relative Poverty Line									
Squared Poverty Gap									
PERFORMANCE FACTORS		None		Design		RR_pp		RR_ph	
True Bias		0.0004		0.0001		0.0013		0.0019	
True MSE		0.0001		0.0001		0.0001		0.0001	
Dev_{bias}		-0.0001 (0.00009)		0.0002 (0.00010)		-0.0010 (0.00011)		-0.0019 (0.00014)	
Dev_{MSE}		0.0000 (0.00000)		0.0000 (0.00000)		0.0001 (0.00000)		0.0001 (0.00000)	
% Rel. Bias: $\widehat{var}(\hat{\theta})$		11.3309		13.2280		110.5743		112.5099	
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$		11.1838		13.4252		105.0235		100.2959	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.4948	0.433	0.4742	0.4639	0.5052	0.4124	0.5258	0.3814
	Perc	0.0412	0.0412	0.0515	0.0619	0	0.0103	0.0103	0.0103
	Boott	0.0309	0.1134	0.4845	0.4639	0.5052	0.4021	0.5361	0.3814
Length:	Std	0.0013		0.0014		0.0019		0.0019	
	Perc	0.0251		0.0258		0.0357		0.0356	
	Boott	0.0218		0.001		0.0019		0.0019	
Std Length:	Std	0.0464		0.0468		0.0631		0.0622	
	Perc	0.8677		0.8735		1.1699		1.1492	
	Boott	0.7528		0.0346		0.0628		0.0616	
Shape:	Std	1		1		1		1	
	Perc	1.176		1.1847		1.2508		1.1981	
	Boott	1.0288		1.199		1.246		1.3077	

Table 7.16: Summary of performance measures calculated for squared poverty gap based on the relative poverty line

No weighting achieved the best results for the estimation of the squared poverty gap index. The standard errors shown are very close to zero indicating that also

here the bootstrap method appears to be reliable. The percentage relative biases under no weighting and design weight are similar and not very big, while those under RR_pp and RR_ph are once again quite large. The best non-coverage was achieved with the percentile confidence interval and its non-coverages seems to be improved slightly by weighting. The total non-coverage closest to 0.05 was achieved under RR_ph with approximately 0.02 total non-coverage. Slight over-estimation seemed to occur here. Again the bootstrap- t confidence interval achieved a pleasing result under no weighting, but this seemed to be worsened by the weighting methods. This same pattern can be seen in its lengths under the various weighting methods. The standard (asymptotic) confidence interval did not give good results which can be verified by its short lengths.

Poverty Measures									
Relative Poverty Line									
At-risk-of-poverty, \hat{I}_1									
PERFORMANCE FACTORS		None		Design		RR_pp		RR_ph	
True Bias		-0.0021		-0.0028		0.0013		0.0038	
True MSE		0.0005		0.0005		0.0006		0.0006	
Dev_{bias}		0.002 (0.00067)		0.0034 (0.00060)		-0.002 (0.00082)		-0.0057 (0.00091)	
Dev_{MSE}		0.0001 (0.00002)		0.0001 (0.00002)		0.0008 (0.00003)		0.0008 (0.00003)	
% Rel. Bias: $\widehat{var}(\hat{\theta})$		12.0484		14.6144		124.6887		135.1289	
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$		11.1082		13.0966		124.1853		129.9379	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.4845	0.433	0.4639	0.4742	0.5567	0.3505	0.6082	0.2887
	Perc	0.0515	0.0722	0.0412	0.0722	0	0.0103	0	0.0103
	Boott	0.1134	0.1753	0.4845	0.4742	0.567	0.3711	0.6392	0.299
Length:	Std	0.0041		0.0042		0.0063		0.0062	
	Perc	0.0769		0.0787		0.1171		0.1155	
	Boott	0.0562		0.0018		0.004		0.0038	
Std Length:	Std	0.045		0.0457		0.0649		0.0653	
	Perc	0.8461		0.8572		1.2159		1.2226	
	Boott	0.6182		0.0196		0.0413		0.0406	
Shape:	Std	1		1		1		1	
	Perc	1.0714		1.0892		0.9743		0.9405	
	Boott	1.132		1.2232		1.4353		1.5108	

Table 7.17: Summary of performance measures calculated for AROP based on the relative poverty line

Similar results were obtained in table 7.17 to what was obtained in table 7.16.

Conclusions: Poverty Measures

The poverty measures were calculated based on two poverty lines, namely the absolute and the relative poverty lines. The best results based on the absolute poverty line were achieved mostly under no weighting, but also for design weight and RR_ph. The relative poverty line achieved best results mainly under no weighting. It was seen throughout the tables that bootstrap resampling seems to be a reliable estimation method. This follows from the small Dev_{bias} and Dev_{MSE} values, the difference between the bootstrap estimated bias/MSE and the true bias/MSE, along with their small standard errors. It seems as if the appropriate confidence interval to use for the poverty measures, is the percentile method. Although some occasional over-estimation occurred, it gave non-coverage probabilities closest to the desired significance level, $\alpha = 0.05$. Along with its good coverage, it also resulted in interval lengths that were quite short.

Inequality Measures									
Percy									
Gini Coefficient									
PERFORMANCE FACTORS		None		Design		RR_pp		RR_ph	
True Bias		-0.0152		-0.0149		0.0002		0.0014	
True MSE		0.0007		0.0007		0.0005		0.0005	
Dev_{bias}		0.014 (0.00015)		0.0138 (0.00014)		-0.0013 (0.00027)		-0.0023 (0.00027)	
Dev_{MSE}		-0.0002 (0.00004)		-0.0002 (0.00003)		0.0002 (0.00006)		0.0002 (0.00005)	
% Rel. Bias: $\widehat{var}(\hat{\theta})$		-0.3156		0.9263		30.0275		29.4689	
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$		-32.9948		-32.056		30.2253		29.0974	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.2165	0.7113	0.2268	0.7216	0.4433	0.4742	0.4845	0.4536
	Perc	0	0.3505	0	0.3402	0.0103	0.0928	0.0103	0.0722
	Boott	0.0412	0.3505	0.0412	0.3505	0.0309	0.0825	0.0309	0.1031
Length:	Std	0.0036		0.0036		0.0044		0.0043	
	Perc	0.065		0.0651		0.0812		0.081	
	Boott	0.0595		0.0588		0.097		0.0947	
Std Length:	Std	0.0345		0.035		0.0479		0.0478	
	Perc	0.628		0.6377		0.8906		0.8923	
	Boott	0.5743		0.5759		1.0639		1.0441	
Shape:	Std	1		1		1		1	
	Perc	0.7387		0.7412		0.9391		0.9451	
	Boott	1.8226		1.8037		1.4345		1.4079	

Table 7.18: Summary of performance measures calculated for Gini coefficient based on percy

Table 7.18 contains varied results. The first section of the table, containing the performance measures of the bootstrap estimation of Gini, shows that RR_pp achieved the best results. The standard errors are all very small, indicative of the reliability of the bootstrap resampling method. The smallest percentage relative bias of $\widehat{var}(\hat{\theta})$ was achieved with no weighting, but a similar percentage was achieved under design weight. Other percentages were quite similar with the smallest percentage relative bias of $\widehat{MSE}(\hat{\theta})$ being achieved by RR_ph. The best non-coverage was achieved by the percentile confidence interval that shows significant improvement of its non-coverage probabilities through the application of the weighting methods. Its best total non-coverage was achieved under RR_ph, namely approximately 0.08. The bootstrap- t confidence interval also showed improvement with the weighting

methods, but the standard (asymptotic) interval once again performed quite poorly. This coincides with it having the shortest length of the three confidence intervals. Percentile and Bootstrap- t confidence intervals show similar lengths with no significant increase in the lengths under the different weighting methods.

Inequality Measures									
Percy									
Mean Log Deviation									
PERFORMANCE FACTORS		None		Design		RR_pp		RR_ph	
True Bias		-0.0558		-0.0559		-0.0009		0.0051	
True MSE		0.0079		0.0078		0.0062		0.0062	
Dev_{bias}		0.0523 (0.00034)		0.0523 (0.00034)		-0.0024 (0.00067)		-0.0082 (0.00067)	
Dev_{MSE}		-0.0031 (0.00036)		-0.0031 (0.00034)		0.0023 (0.00075)		0.0022 (0.00068)	
% Rel. Bias: $\widehat{var}(\hat{\theta})$		-0.236		1.1919		36.181		35.6205	
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$		-39.3629		-39.2431		36.3348		35.2133	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.1959	0.7526	0.1856	0.7423	0.3814	0.5052	0.4536	0.4536
	Perc	0	0.3608	0	0.3711	0.0103	0.1134	0.0206	0.0825
	Boott	0.0412	0.3299	0.0515	0.3299	0.0309	0.0825	0.0412	0.0825
Length:	Std	0.0115		0.0115		0.015		0.015	
	Perc	0.2102		0.2104		0.2784		0.2798	
	Boott	0.1943		0.1922		0.3307		0.3245	
Std Length:	Std	0.033		0.0332		0.0486		0.0486	
	Perc	0.6025		0.6084		0.8989		0.9056	
	Boott	0.5568		0.5555		1.0678		1.0501	
Shape:	Std	1		1		1		1	
	Perc	0.795		0.798		1.0123		1.0112	
	Boott	2.0338		2.0182		1.7158		1.7128	

Table 7.19: Summary of performance measures calculated for MLD based on percy

Here too varied results were obtained. The performance measures of the bootstrap estimation of MLD achieved best results under both RR_pp and RR_ph weighting. Very small standard errors were obtained showing not much difference between the weighting methods. All standard errors were close to zero. The reliability of the bootstrap method thus seems reasonable. The percentage relative bias of a variance estimator was the smallest under no weighting, with design weight achieving not much larger percentage. The percentage relative biases of MSE achieved

similar percentages over all weighting methods, but was the smallest under RR_ph. It seems as if the percentile and bootstrap- t confidence intervals achieved quite similar total non-coverage probabilities, with the percentile interval being only slightly better. Their total non-coverage was improved by weighting, but the smallest total non-coverage, under RR_ph, was approximately 0.1. This is double the desired significance level, 0.05. The similar lengths of these two confidence intervals confirm their similar non-coverage probabilities. The standard (asymptotic) confidence interval performed poorly once again, confirmed by its very short length.

Inequality Measures									
Percy									
Theil T									
PERFORMANCE FACTORS		None		Design		RR_pp		RR_ph	
True Bias		-0.0565		-0.0536		-0.0052		-0.0033	
True MSE		0.0238		0.0229		0.0269		0.0265	
Dev_{bias}		0.0462 (0.00111)		0.0434 (0.00106)		-0.012 (0.00223)		-0.0121 (0.00209)	
Dev_{MSE}		-0.0073 (0.00226)		-0.007 (0.00211)		-0.0008 (0.00334)		-0.001 (0.00323)	
% Rel. Bias: $\widehat{var}(\hat{\theta})$		-20.5349		-20.8229		-4.081		-4.6368	
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$		-30.7619		-30.2924		-3.0782		-3.7813	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.299	0.6804	0.3093	0.6701	0.4124	0.5361	0.3711	0.5258
	Perc	0	0.4227	0	0.4124	0.0103	0.2577	0.0103	0.2371
	Boott	0.1134	0.4021	0.1031	0.3711	0.0206	0.1959	0.0206	0.1856
Length:	Std	0.0189		0.0189		0.024		0.0238	
	Perc	0.3277		0.3281		0.4397		0.4394	
	Boott	0.4855		0.4718		0.8231		0.7993	
Std Length:	Std	0.0313		0.0318		0.0373		0.0373	
	Perc	0.5419		0.5525		0.6838		0.6885	
	Boott	0.8029		0.7945		1.2802		1.2525	
Shape:	Std	1		1		1		1	
	Perc	0.7022		0.7056		0.9014		0.9213	
	Boott	3.2267		3.1918		2.411		2.3719	

Table 7.20: Summary of performance measures calculated for theil T based on percy

The performance measures of the bootstrap estimation of Theil T shows that RR_pp achieved the best results. Although it has slightly larger standard errors than no weighting, all standard errors are close to zero indicating that the differ-

ence is negligible. Both RR_pp and RR_ph achieved much smaller percentage relative biases than no weight and design weight. The percentile confidence interval performed best under no weighting and design weight, while the bootstrap- t confidence interval performed best under RR_pp and RR_ph. These also resulted in the smallest total non-coverage probabilities, these probabilities still being significantly different from 0.05. Perhaps these confidence intervals were too conservative for the Gini, as can be seen from their lengths. It can also be concluded from the interval lengths that the poor performance of the standard (asymptotic) confidence interval is expected.

Inequality Measures									
Equivalized Income									
Gini Coefficient									
PERFORMANCE FACTORS		None		Design		RR_pp		RR_ph	
True Bias		-0.0154		-0.0155		0.0003		0.0001	
True MSE		0.0007		0.0007		0.0006		0.0006	
Dev_{bias}		0.012 (0.00022)		0.0121 (0.00021)		-0.0081 (0.00050)		-0.0073 (0.00046)	
Dev_{MSE}		0 (0.00006)		0 (0.00006)		0.0009 (0.00011)		0.0009 (0.00010)	
% Rel. Bias: $\widehat{var}(\hat{\theta})$		52.7593		57.0074		138.0991		140.9894	
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$		4.2219		5.4392		148.0364		149.9438	
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.1959	0.7629	0.1649	0.7732	0.4021	0.4948	0.4124	0.4639
	Perc	0	0.2371	0	0.2165	0	0.0309	0	0.0309
	Boott	0	0.2062	0.2165	0.7732	0.4742	0.5155	0.4639	0.5155
Length:	Std	0.0045		0.0045		0.0063		0.0063	
	Perc	0.0847		0.0846		0.1185		0.1177	
	Boott	0.092		0.0009		0.0017		0.0017	
Std Length:	Std	0.0429		0.0434		0.0656		0.0662	
	Perc	0.8036		0.8117		1.2304		1.2429	
	Boott	0.8736		0.0089		0.018		0.018	
Shape:	Std	1		1		1		1	
	Perc	0.798		0.7929		0.8605		0.8703	
	Boott	1.4157		1.3858		1.3168		1.3099	

Table 7.21: Summary of performance measures calculated for Gini based on equivalized income

Table 7.21 shows varied results. The performance measures of the bootstrap

estimation of Gini show “equal effectiveness” by using no weighting of RR_ph. The standard errors remain close to zero, with the standard errors under no weight and design weight slightly smaller. This difference between the standard errors of weighting and not weighting can still be ignored, since all standard errors were very close to zero. Thus, bootstrap resampling seems to be reliable for the estimation of Gini based on equivalized income. The percentage relative bias of a variance estimator was larger than the percentage bias of a MSE estimator under no weight and design weight. Both RR_pp and RR_ph resulted in very large percentage relative biases in comparison to no weight and design weight. The total non-coverage probabilities of the percentile confidence were smallest and significantly improved by both RR_pp and RR_ph. These weighting methods resulted in total non-coverage of approximately 0.03, very close to the desired significance level of 0.05. Both the standard (asymptotic) and the bootstrap- t confidence intervals performed quite poorly. The reason for this can be concluded from their lengths that were consistently much shorter than that of the percentile interval.

Inequality Measures									
Equivalized Income									
Mean Log Deviation									
PERFORMANCE FACTORS	None	Design		RR_pp		RR_ph			
True Bias	-0.0488	-0.0498		-0.0001		-0.0002			
True MSE	0.0065	0.0064		0.0057		0.0055			
Dev_{bias}	0.044 (0.00045)	0.0448 (0.00045)		-0.0128 (0.00107)		-0.0115 (0.00098)			
Dev_{MSE}	0 (0.00054)	-0.0001 (0.00051)		0.0097 (0.00139)		0.0095 (0.00125)			
% Rel. Bias: $\widehat{var}(\hat{\theta})$	57.5035	60.9525		168.0023		170.0289			
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$	-0.3409	-1.093		170.9583		172.5482			
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.1856	0.7732	0.1649	0.7835	0.3711	0.5052	0.3918	0.4742
	Perc	0	0.2474	0	0.2268	0	0.0206	0	0.0412
	Boott	0	0.2165	0.1856	0.7835	0.4227	0.5052	0.433	0.5052
Length:	Std	0.0132		0.0132		0.0202		0.02	
	Perc	0.2473		0.2467		0.3707		0.3675	
	Boott	0.278		0.0067		0.0141		0.0137	
Std Length:	Std	0.0419		0.042		0.0683		0.0689	
	Perc	0.7856		0.7868		1.2543		1.2667	
	Boott	0.8828		0.0214		0.0476		0.0472	
Shape:	Std	1		1		1		1	
	Perc	0.9476		0.9419		1.0901		1.0996	
	Boott	1.4309		1.3278		1.2393		1.222	

Table 7.22: Summary of performance measures calculated for MLD based on equivalized income

Similar conclusions can be drawn than those of table 7.21.

Inequality Measures									
Equivalized Income									
Theil T									
PERFORMANCE FACTORS	None	Design		RR_pp		RR_ph			
True Bias	-0.0594	-0.0582		-0.009		-0.011			
True MSE	0.0235	0.0227		0.0278		0.0265			
Dev_{bias}	0.045 (0.00169)	0.0439 (0.00162)		-0.0257 (0.00394)		-0.0218 (0.00353)			
Dev_{MSE}	-0.0025 (0.00312)	-0.0021 (0.00298)		0.0119 (0.00514)		0.012 (0.00478)			
% Rel. Bias: $\widehat{var}(\hat{\theta})$	4.1706	5.4509		38.9537		41.8411			
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$	-10.5545	-9.3895		42.8978		45.2479			
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.2371	0.7216	0.2474	0.7216	0.3402	0.5979	0.3093	0.6186
	Perc	0	0.3711	0	0.3711	0	0.1753	0	0.1546
	Boott	0	0.299	0.2577	0.732	0.3402	0.6082	0.3299	0.6289
Length:	Std	0.0211		0.0211		0.0295		0.0292	
	Perc	0.3878		0.3878		0.5463		0.5427	
	Boott	0.6535		0.0175		0.0299		0.0291	
Std Length:	Std	0.0351		0.0357		0.0451		0.0458	
	Perc	0.6448		0.6573		0.8363		0.851	
	Boott	1.0864		0.0296		0.0457		0.0457	
Shape:	Std	1		1		1		1	
	Perc	0.8793		0.8792		1.0239		1.0357	
	Boott	2.042		1.5591		1.504		1.4772	

Table 7.23: Summary of performance measures calculated for theil T based on equivalized income

Table 7.23 shows that design weight achieved the best results for the estimation of Theil T. Slightly larger standard errors were achieved than in the previous tables, but still quite small. None of these standard errors suggest that the bootstrap method is unreliable in the estimation of Theil T. The percentage relative biases were smallest under no weight and design weight, but both RR_pp and RR_ph resulted in smaller percentages than in some of the previous tables. Once again the percentile confidence interval achieved the best non-coverage probabilities and was significantly improved by the weighting methods. However, its best total non-coverage is approximately 0.15, significantly different from 0.05. The standard (asymptotic) and bootstrap- t confidence intervals performed quite badly. Their very short lengths confirm why this is the case, while the percentile confidence interval lengths are slightly longer,

hence achieving better non-coverage.

Inequality Measures									
Equivalized Income									
Quintile Share Ratio, \hat{I}_2									
PERFORMANCE FACTORS	None	Design		RR_pp		RR_ph			
True Bias	-2.8821	-2.9957		-0.0686		0.0262			
True MSE	19.1137	19.0101		16.5867		16.0542			
Dev_{bias}	2.5803 (0.03680)	2.6226 (0.03777)		-0.8649 (0.05893)		-0.8969 (0.05861)			
Dev_{MSE}	-0.4619 (1.24157)	-0.6144 (1.15383)		43.7195 (6.07610)		42.1625 (5.23413)			
% Rel. Bias: $\widehat{var}(\hat{\theta})$	71.7475	81.9102		258.4298		257.92			
% Rel. Bias: $\widehat{MSE}(\hat{\theta})$	-2.4163	-3.2319		263.5825		262.6266			
		Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
NCP	Std	0.1649	0.7732	1	0	1	0	1	0
	Perc	0	0.2165	1	0	1	0	1	0
	Boott	0	0.2577	1	0	1	0	1	0
Length:	Std	0.7211	0.7178	1.2497	1.2402				
	Perc	13.4675	13.3859	22.1215	21.9035				
	Boott	12.1896	0.5483	1.3961	1.3672				
Std Length:	Std	0.0421	0.042	0.0783	0.079				
	Perc	0.7858	0.7832	1.3857	1.3946				
	Boott	0.7113	0.0321	0.0875	0.0871				
Shape:	Std	1	1	1	1				
	Perc	1.0572	1.0221	1.152	1.1569				
	Boott	1.4479	1.3625	1.3467	1.3342				

Table 7.24: Summary of performance measures calculated for QSR based on equivalized income

From table 7.24 it can be concluded that design weight achieved best results in the estimation of QSR. The standard errors are slightly larger in table 7.24 than in the previous tables. In comparison to the previous standard errors this could indicate that the bootstrap method is perhaps not ideal for estimating the QSR, but the standard errors are still quite small. Further research would have to be done in this regard. Varied results were obtained for percentage relative biases. RR_pp and RR_ph resulted in very large percentages while no weight and design weight achieved much smaller percentage relative biases. The percentile confidence interval still achieved the smallest total non-coverage probability, but here the weighting

methods did not improve these non-coverage probabilities and hence the best result was obtained under no weighting. This being said, all confidence intervals did not seem to work well for the QSR, as can be seen from the very large interval lengths. Further research would also have to be done in this regard.

Conclusions: Inequality Measures

The inequality measures were based on two different income variables, namely percy and equivalized income. In the case of percy, it seems as if RR_pp achieved the best results, while in the case of equivalized income, there were mixed results between no weight, design weight and RR_ph. Except for QSR, Dev_{bias} and Dev_{MSE} resulted in very small values, along with their standard errors being close to zero. This was a common occurrence over all weighting and not weighting and thus it must be concluded that the bootstrap method is reliable for the estimation of the inequality measures, with the exception of QSR. The percentile confidence interval seemed to be best suited for the inequality measures, although it also contained varied results in terms of its total non-coverage probability. In contrast to the poverty measures, the percentile interval showed signs of under-estimation, along with some over-estimation. It also achieved very short lengths along with its pleasing non-coverage.

Overall Conclusions

In general, it is possible that the varied results could be attributed to the fact that only 100 bootstrap populations were used from the surrogate population. In fact, the surrogate population's use is perhaps questionable in itself: it is an extremely skew population (at least in terms of the income distribution) that could require a larger number of bootstrap resamples to be taken. The use of the relative poverty line introduced additional variability, the effect which was not studied, but compensated for by recalculating it for every surrogate population, as well as for every bootstrap resample. These issues should be addressed in further study.

However mixed the results may be, it can be concluded that the estimation of certain welfare indices seems to be improved by weighting. A question that comes to mind is whether other auxiliary variables (not necessarily in the current data set) should have been used in the calibration. Variable selection techniques could be

implemented to choose the “best set” of auxiliary variables. This is another issue for further study.

The reliability of the resampling methods seems to be quite positive. The difference between the mean estimated bias/MSE and the true bias/MSE shows promising results for most estimators. The same can be said for the percentage relative bias calculations. Perhaps the mixed results relate more to the weighting and the type of poverty line used, rather than the resampling methods used (refer section 2.4.4). As for the confidence intervals, the percentile interval performed well with some promise being shown by the bootstrap- t interval. Perhaps the bootstrap- t confidence interval could be improved by using the bootstrap resampling method for the estimation of the variances of the bootstrap estimators instead of the jackknife method, as was the case in this study. This will also be addressed in further study.

It should be emphasized once again that this was an extremely complex situation for which estimation was done and confidence intervals were calculated. It could thus be expected that the results would be quite varied. Similar variation in results also occurred in literature [14].

It could thus be concluded that weighting seems to be worthwhile for certain estimators and that bootstrap resampling can in many cases be considered a reliable estimation method for complicated data structures and estimators.

Chapter 8

Conclusions and Further Research

The purpose of this study was to obtain confidence intervals for the estimators of poverty and inequality indices under a complex sampling design. It started off with a discussion on sampling techniques in general and then focused on complex sampling. An important part of complex sampling is weighting and hence attention was paid to its construction during different stages of the sample design as well as its use in estimation.

In order to obtain the estimators of the welfare indices, followed by confidence intervals for them, resampling methods were used and discussed in the study. Although other resampling methods do exist, such as linearization and balanced repeated replication (BRR), it was decided to focus on the jackknife and bootstrap resampling methods, due to their pleasing properties and “easy” application to any data set. The bootstrap method was used for the estimation of the variances, biases and mean squared errors of the estimators and the jackknife method was used for the estimation of the variances of the bootstrap estimators. This was done primarily because using the bootstrap for the latter involves a nested bootstrap which becomes computationally expensive. From these calculations the standard (based on asymptotic normality), percentile and bootstrap- t confidence intervals were constructed for the estimators of the welfare indices. The resampling methods as well as the confidence intervals and their construction, were discussed in detail.

The welfare indices used in this study were defined and their estimation in a complex survey design was discussed in detail. Included in this discussion was Atkinson’s inequality measures, but these were not calculated in the simulation study. The study of these indices will be considered in further research.

In order to carry out the study based on a realistic South African data set, the Income and Expenditure survey of 2005/2006 was used as a surrogate population. This survey data set is based on a complex survey procedure and a Monte Carlo simulation was carried out from this population to form 100 bootstrap populations, based on complex sampling schemes similar to that of the surrogate population. These bootstrap populations were then subjected to different estimation procedures in order to estimate the quantities of interest and to construct the confidence intervals for them.

To determine the performance of weighting versus non weighting, as well as the reliability of using resampling methods in complex survey data, different measures of accuracy were calculated in the simulation study. The main conclusions of the simulation study were the following:

- In most large surveys, units have unequal inclusion probabilities and problems such as non-response and non-coverage occur. To ensure that the survey represents the target population as closely as possible, it is necessary to address these issues. This is one of the main reasons for weighting. The design weight of a household is constructed at different stages, as discussed in section 2.4, where it is adjusted to compensate for non-response and non-coverage. Calibration and integrated weighting are methods that adjust the design weight through the use of known auxiliary variables. These could either be person variables or person and household variables simultaneously. Calibration results in a different weight for each household member while integrated weighting results in a single set of weights that could be used for both person and household estimation. Thus, weighting seems like a reasonable method to use in analysis. Calibration weighting makes use of a distance function and in this study the raking ratio method was used. This resulted in RR_pp, raking ratio making use of only person auxiliary variables, and RR_ph, raking ratio making use of person and household auxiliary variables. Mixed results were obtained with regard to the no weighting and weighting. Weighting achieved improvement in many cases but not all. It especially improved the non-coverage of the percentile confidence interval. Further research should be done to determine how weighting should be applied in estimating welfare indices - see below.
- Two poverty lines were used in the study. The absolute poverty line is tra-

ditionally used, also in South Africa. The relative poverty line is a newer proposal and is of course random, since it depends on the sample data. It is the preferred method used in the European Laeken indices. For our data, some differences occurred between the results obtained from the poverty lines.

- The performance of the resampling methods were judged in terms of
 - Dev_{bias} and its standard error;
 - Dev_{MSE} and its standard error;
 - Percentage relative bias of a variance estimator; and
 - Percentage relative bias of a MSE estimator.

All welfare indices, except QSR, showed very small deviances between the bootstrap estimated bias/MSE and the true bias/MSE. The standard errors that accompanied these measures were all very close to zero, again with the exception of QSR. The percentage relative biases, on the other hand, were not always very good. Mostly they seemed acceptable for no weight and design weight with occasionally being acceptable for either RR_pp or RR_ph, with the latter two mainly resulting in larger percentages in comparison to no weight and design weight. The percentage relative biases seemed more acceptable for measures based on the absolute poverty line (percy) than for those based on the relative poverty line (equivalized income). This calls for further research on this aspect.

- The performance of the confidence intervals can mainly be summarised in terms of their coverage probability and length. These two measures are closely related. A low coverage probability implies a short interval length, and vice-versa. The data clearly followed this path as seen in the tables in section 7.3. It could be concluded that the percentile confidence interval was better suited for this data than the standard (asymptotic) or the bootstrap- t confidence intervals based on total non-coverage probability and length. It mostly resulted in total non-coverage closer to the desired significance level, $\alpha = 0.05$, along with fairly short interval lengths. Further research will be done on improving the confidence intervals.

Based on these conclusions, the following areas for further research are mentioned:

1. The IES data set turned out to be quite a strange data set, but at the time of this study it was the only data set that was available. These techniques have to be tested on other data sets. Further research would make use of “better” data sets.
2. Investigate if increasing the number of bootstrap populations as well as the number of bootstrap resamples taken from each surrogate population, makes a difference to the results.
3. Study the effect of the relative poverty line, since varied results were obtained for estimators based on it.
4. The use of calibration and integrated weighting does show potential in many cases. Given the mixed results obtained for `RR_pp` and `RR_ph`, it should be determined whether the correct auxiliary variables were used in their construction, or if other variables, given their availability, should have been considered. Variable selection techniques could be implemented to determine the “best set” of auxiliary variables to use in calibration and integrated weighting. Further research should be done in this regard.
5. Bootstrap methods can be used for estimation of the quantities of interest since no evidence revealed that it did not result in reliable estimates. Further research should however be done on the application of the bootstrap method for the estimation of QSR.
6. Using a combination of the bootstrap and jackknife is quite efficient for the construction of confidence intervals for welfare indices using complex survey data. Further research that should be done here is the application of the nested bootstrap approach for the construction of the bootstrap- t confidence interval as well as the application of other bootstrap confidence intervals in complex surveys, such as the BCa confidence interval and the ABC confidence interval.

Bibliography

- [1] Statistics South Africa. Income and expenditure of households 2005/2006: Statistical release. Technical report, Statistics South Africa, 2008.
- [2] A.B. Atkinson. On the measurement of inequality. *Journal of Economic Theory*, 2:244–263, 1970.
- [3] J.G. Booth and S. Sarkar. Monte carlo approximation of bootstrap variances. *The American Statistician*, 52:354–357, 1998.
- [4] A. DasGupta. *Asymptotic Theory of Statistics and Probability*. Springer, 2008.
- [5] T. De Wet and A. Neethling. Accuracy of poverty and inequality analyses under complex sampling. Unpublished Manuscript, 2008.
- [6] A. Demnati and J.N.K. Rao. Linearization variance estimators for survey data. *Statistics Canada*, 30:17–26, 2004.
- [7] J. C. Deville and C. E. Särndal. Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382, 1992.
- [8] Särndal C. E. Deville, J. C. and O. Sautory. Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88:1013–1020, 1993.
- [9] T. J. DiCiccio and B. Efron. Bootstrap confidence intervals. *Statistical Science*, 11:189–212, 1996.
- [10] B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, 1993.

- [11] J. Haughton et. al. *Introduction to Poverty Analysis*. World Bank Institute, August 2005.
- [12] Unit E-2: Living Conditions EuroStat, Directorate E: Social Statistics. Laeken indicators: Detailed calculation methodology. Technical report, European Commission, April 2003.
- [13] K. Knight. *Mathematical Statistics*. Chapman & Hall/CRC, 2000.
- [14] M. Kovacevic and W. Yung. Variance estimation for measures of income inequality and polarization: An empirical study. *Statistics Canada*, 23:41–52, 1997.
- [15] Rao J.N.K. Kovar, J.G. and C.F.J. Wu. Bootstrap and other methods to measure errors in survey estimates. *The Canadian Journal of Statistics*, 16:25–45, 1988.
- [16] E. Leiten and I. Traat. Variance of laeken indicators in complex surveys. Technical report, Statistical Office of Estonia, 2004.
- [17] G. Lemaitre and J. Dufour. An integrated method for weighting persons and families. *Survey Methodology*, 13:199–207, 1987.
- [18] S.L. Lohr. *Sampling: Design and Analysis*. Brooks/Cole Publishing Company, 1999.
- [19] A. Neethling. Calibration and integrated weighting. Course notes, University of Stellenbosch, 2009.
- [20] A. Neethling. *Calibration and Integrated Weighting in Sample Surveys*. PhD thesis, Faculty of Science, University of Witwatersrand. Additional Notes.
- [21] A. Neethling. Complex sampling. Course notes, University of Stellenbosch, 2009.
- [22] A. Neethling and J.S. Galpin. Weighting of household survey data: A comparison of various calibration, integrated and cosmetic estimators. *The South African Statistical Journal*, 40:123–150, 2006.

- [23] Department of Economic and Social Affairs. Designing household survey samples: Practical guidelines. Technical Report Series F No. 98, United Nations, 2008.
- [24] H. Pettersson. The design of a master sampling frame and master sample for surveys in developing countries. Technical report, Statistics Sweden, 2005.
- [25] J.N.K. Rao and C.F.J. Wu. Methods for standard errors and confidence intervals from sample survey data. *Bulletin of the International Statistical Institute*, 1987.
- [26] Wu C.F.J. Rao, J.N.K. and K. Yue. Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18:209–217, 1992.
- [27] K.F. Rust and J.N.K. Rao. Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5:283–310, 1996.
- [28] R. F. G. Williams and D. P. Doessel. Measuring inequality: Tools and an illustration. *International Journal for Equity in Health*, 5, 2006.